TESE DE DOUTORAMENTO

ANALYSIS OF CHARGED PARTICLE PRODUCTION IN PROTON-NUCLEUS AND PROTON-PROTON COLLISIONS AT THE LHCB EXPERIMENT

Óscar Boente García

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN FÍSICA NUCLEAR E DE PARTÍCULAS

SANTIAGO DE COMPOSTELA

2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DO AUTOR/A DA TESE

D./Dna. **Óscar Boente García**

Título da tese: **Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4) A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de Compostela**, **22 de Xuño de 2021**.

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR/TITOR DA TESE

D./Dna. **Abraham Antonio Gallas Torreira**

En condición de: **Titor/a e director/a**

Título da tese: **Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment**

INFORMA:

Que a presente tese, correspóndese co traballo realizado por D/Dna Óscar Boente García, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director/titor desta non incorre nas causas de abstención establecidas na Lei 40/2015.

En **Santiago de Compostela**, **22 de Xuño de 2021**

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR/TITOR DA TESE

D./Dna. **Cibrán Santamarina Ríos**

En condición de: **Director/a**

Título da tese: **Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment**

INFORMA:

Que a presente tese, correspóndese co traballo realizado por D/Dna Óscar Boente García, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director/titor desta non incorre nas causas de abstención establecidas na Lei 40/2015.

En **Santiago de Compostela**, **22 de Xuño de 2021**

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Acknowledgements

Firstly, I would like to thank my supervisors Abraham and Cibrán for their great support during these years that led me to complete this thesis. I am sure that it would not have been possible without their advise and help at the tough times. Thanks to Álvaro, who started this analysis and lent me the products of his hard work, and helped me to complete my Master Thesis and during the beginning of my PhD. Also thanks to Ricardo, who has helped me enormously with the final push to this analysis and with the paper preparation. I am going to miss our lively discussions regarding the paper, but I will keep everything that I learned.

Thanks to professor Giulia Manca for inviting me to the Cagliari IFT meetings, where I have learned and shared many of the problems I had with the analysis, and then, for welcoming me in my short stay at Cagliari University. With her I want to thank to all her team for their help and patience: Benjamin, Samuel, Roman, Albert, Shanzhen and Jiayin. I think I learned there a remarkable lesson for a scientist, which is how important is to communicate properly the issues in your research to get the proper help and collaborate with other scientists.

Also I would like to thank Frédéric for welcoming me at during my short stay at the Laboratoire Leprince-Ringuet at Palaiseau. Also, thanks to the rest of the members of the LHCb group in LLR: Emilie, Benjamin, Felipe and Vladik. I really enjoyed working together and I really appreciate your support. Special thanks to Ben, for being a friend and making easier to live with the pandemic in those months. And to Vladik, for his implication in the *pp* luminosity measurement which is a fundamental input for this analysis.

My gratitude to all people of the VELO group, in particular Victor Coco, for his support during my stay at CERN. Also thanks to the USC VELO team, specially to Eliseo, who showed me the basics of electronics and helped me during my short period in the laboratory measuring sensors for my bachelor thesis.

I would like to continue thanking all the coverners that were leading the IFT group at LHCb during these years: Giacomo, Francesco, Frédéric, Michael, Benjamin, Yanxi, Mat and Jiayin. I was always welcomed at the IFT meetings and I felt that my work was valued. More in general, it is important to remember other people who I discussed the analysis with, like Hans, since his points of view contributed to the improvement of my work. To the LHCb reviewers of the analysis, Xianglei Zhu and Andrea Merli, thanks for their interesting and useful comments, and for making possible a detailed review in a record time. Then, thanks to the members of the LHCb Editorial Board, specially Sergey Barsuk, but also Philip Ilten and Antimo Palano, for their comments to the paper draft. I am also very grateful to Elena and Néstor, part of the IGFAE QCD phenomenology group, for the interesting discussions that we had all over these years and their help to understand better the more theoretical aspects of this thesis.

I thank the financial support provided by the "Programa de Formación de Profesorado

Universitario" of the "Ministerio de Universidades", which founded my contract at USC and my stays at Cagliari and Palaiseau.

No lado máis persoal, comezar agradecendo aos meus compañeiros de despacho e traballo, xa amigos e acompañantes de dramas e comedias a este punto. Primeiro a María, que sempre foi un enorme apoio e a mellor mentora posible para navegar no mundo do doutoramento, a investigación e a burocracia. A Bea, Ale e Julián, por ser os meus compañeiros de aventuras polo mundo. Tamén me gustaría darlles forza aos novos que entran no grupo, Imanol e Sara, que teñan ganas para enfrontarse a este reto. Agradecer tamén a todo o resto dos meus compañeiros do grupo de LHCb, tanto estudantes de doutorando, como seniors (Toño, Juan,...), por crear un ambiente de traballo agradable. Mencionar, máis en xeral, a todas as amizades que atopei polo camiño e que fixeron todo máis levadío. Tamén me gustaría recordar aos meus amigos, que me axudastes a equilibrar a balanza e non afogarme (tanto) co traballo. Sara, María, Vero, Jenni, Carlos, Domin, Jessica..., e todos os meus amigos, tanto de Santiago como de Ponteareas. Por último a toda a miña familia, pero sobre todo aos meus pais Puri e Manolo e a miña irmá Estela, por ser un grande apoio para continuar.

Abstract

This thesis presents the measurement of prompt charged particle production in proton-lead and proton-proton collisions at the centre-of-mass energy of $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ in the LHCb experiment at CERN. Double-differential cross-sections of prompt charged particles are determined in the two systems as a function of pseudorapidity (η) and transverse momentum ($p_{\rm T}$) with respect to the proton beam.

With the measured cross-sections, the nuclear modification factor $R_{p\rm Pb}$ for charged particles is determined covering an acceptance of $-4.8 < \eta < -2.5$ (backward region) and $2.0 < \eta < 4.8$ (forward region), and $0.2 < p_{\rm T} < 8.0\,\text{GeV}/c$. The results show a suppression of charged particle production in proton-lead collisions relative to proton-proton collisions in the forward region and an enhancement in the backward region for $p_{\rm T}$ larger than $1.5\,\text{GeV}/c$. This measurement constrains nuclear PDFs and saturation models at previously unexplored values of the parton fractional momentum down to 10^{-6}.

The thesis has the following structure. In Chap. 1, an **introduction** to the topic is provided. The **objectives** and the **hypotheses** are mentioned there, and are discussed in further detail in Chap. 2. Then, the **methodology** of the analysis is addressed throughout several chapters: starting with the description of the datasets (Chap. 4) and continuing with their preparation and selection (Chap. 5), Chap. 6 addresses the core of this work, which is the description of the analysis techniques to determine the charged particle spectra. The treatment of the systematic uncertainties is tackled in Chap. 7. The **results** of the measurement are presented and discussed in Chap. 8. The final remarks and **conclusions** of this work are provided in Chap. 9. A **summary in Galician** is presented in App. A, and the **bibliography** is included at the end of the document.

Key words: Experimental High Energy Physics, LHCb, Heavy Ions Physics, QCD, Inclusive Prompt Charged Particles, Proton-proton Collision, Proton-Lead Collision, Nuclear Modification Factor, Cold Nuclear Matter

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumo

Esta tese presenta a medida da produción inclusiva de partículas cargadas primarias en colisións protón-chumbo e protón-protón cunha enerxía no centro de masas de $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ no experimento LHCb do CERN. As seccións eficaces diferenciais de partículas cargadas mídense en ámbolos dous sistemas con respecto da pseudorrapidez (η) e do momento transverso (p_{T}), tomando como referencia o sentido do feixe de protóns.

Cos resultados obtidos para as seccións eficaces diferenciais, o factor de modificación nuclear $R_{p\mathrm{Pb}}$ para partículas cargadas primarias determínase en función de η e p_{T} nunha aceptancia de $-4.8 < \eta < -2.5$ (rexión cara diante) e $2.0 < \eta < 4.8$ (rexión cara atrás), e $0.2 < p_{\mathrm{T}} < 8.0\,\mathrm{GeV}/c$. Os resultados asoman unha supresión da produción de partículas cargadas primarias en colisións protón-chumbo en relación con colisións protón-protón escaladas polo número de nucleóns no chumbo na rexión cara diante, mentres que se observa un incremento da produción na rexión cara atrás para p_{T} maiores de $1.5\,\mathrm{GeV}/c$. Esta medida impón límites esixentes ás funcións nucleares de distribución de partóns (nPDF) e aos modelos de saturación en rexións do momento fraccionario dos partóns non exploradas previamente ata 10^{-6}.

A tese ten a seguinte estrutura. No Capítulo 1, realízase unha **introdución** á motivación teórica e identifícanse magnitudes relevantes. Neste capítulo, menciónanse os **obxectivos** e as **hipóteses**, que son discutidas con máis detalle no Capítulo 2. Logo, a **metodoloxía** da análise abórdase en varios capítulos: comezando coa descripción das mostras de datos (Capítulo 4) e continuando coa súa preparación e selección (Capítulo 5), o Capítulo 6 constitúe o núcleo deste traballo, onde se describen as técnicas de análise para determinar o espectro de partículas cargadas primarias. O tratamento das incertezas sistemáticas faise no Capítulo 7. Os **resultados** da medida preséntanse e discútense no Capítulo 8. As derradeiras observacións e **conclusións** deste traballo proporciónanse no Capítulo 9. Un resumo en galego da tese inclúese no Apéndice A, e a **bibliografía** localízase ao final do documento.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

INDEX

1

Introduction

Quantum Chromodynamics (QCD) [1] is the current theory to describe the strong interaction, the force responsible for the existence of composite objects, called hadrons, such as protons and neutrons. QCD provides a solid framework to perform predictions of particle production cross-sections of interactions of such hadrons [2, 3]. However, these predictions are limited to hard interactions, which are those where the scale, given by the transferred momentum, is larger than the QCD parameter Λ_{QCD}. On the contrary, interactions involving small momentum transfer, or soft, are theoretically challenging, complicating the understanding of the underlying processes in hadronic collisions.

One way to study the innermost structure of hadrons, and thus the properties of QCD, is to use particle accelerators to make them collide at ultra-relativistic energies. The Large Hadron Collider (LHC) is the largest hadron accelerator in the world, built at CERN, and performs proton-proton and proton-lead collisions at the highest energies ever reached in a laboratory. The LHCb experiment [4] is located at the LHC in collision point 8. The LHCb spectrometer can detect charged particles produced in the interaction, and determine their kinematics. The acceptance of LHCb is unique at LHC and covers particles produced at low angles with respect to the trajectories of the colliding hadrons.

In these collisions, charged particles like $\pi^{\pm}$, $K^{\pm}$ and $(\overline{p})$ are among the main products and, for the most abundant low momentum charged particles, their production mechanisms are dominated by soft interactions. Currently, first-principles-based predictions of their production cross-sections in high-energy proton-proton and proton-lead collisions are far from achievable, which is a showstopper for research in many areas where hadronic collisions are present. This includes hadron collider physics, where soft charged particle production represents most of the underlying event acting as background to any other pro-

cesses of interest. But also for cosmic ray physics, since the detection of very high-energy cosmic-rays, performed in the Earth surface in facilities such as the Pierre Auger Observatory, relies on accurate modelling of hadronic showers throughout the atmosphere [5–7]. Many Monte-Carlo models try to address this problem but they rely on experimental data to improve their accuracy [8].

The challenge, already formidable for proton-proton collisions, is even more intricate in collisions involving heavy-ions [9]. Under the currently established picture of ion-ion collisions, a quark gluon plasma (QGP) state is formed, where the constituents of hadrons are in a liquid-like configuration. The formation of this state affects the production of charged particles resulting in a suppression of its rate relative to proton-proton [10, 11]. The research of such state relies generally on proton-ion collisions, where such state is in principle not formed in large amounts, as a baseline to interpret the experimental signatures from ion-ion. In proton-lead collisions, charged particle production is affected by the differences in the partonic structure of the lead nuclei with respect to the proton [12]. This modification of the charged particle rate should also manifest in ion-ion collisions but overlaid with QGP effects, therefore it is mandatory to have a deep understanding of the firsts to interpret the phenomenology in both systems correctly. Theoretical models use the hard part of the charged particle spectra, treatable with perturbative techniques, to characterise the nuclear effects in proton-lead collisions, but they need to be validated with experimental data.

This work addresses the aforementioned concerns with a precise measurement of the charged particle production rate in high-energy proton-proton and proton-lead collisions at the LHCb experiment. It is based on the LHCb analysis LHCb-ANA-2020-048 [13], of which the author is the main proponent. The thesis is organised as follows. In Chap. 2, a general outlook of the fundamental theoretical background to understand proton-proton and proton-lead collisions is presented. The overview is focused on those topics most related to the measurement, and towards the end, the targeted physical observables are introduced. Then, in Chap. 3, the LHCb detector is contextualised and described in detail. The rest of the text covers the analysis work, starting with the description of the datasets (Chap. 4) and continuing with their preparation and selection (Chap. 5). Chap. 6 addresses the core of this work, which is the description of the analysis techniques to determine the charged particle spectra. The treatment of the systematic uncertainties is tackled in Chap. 7. The results of the measurement are finally presented and discussed in Chap. 8. The final remarks and conclusions of this work are left for Chap. 9.

2

Theoretical foundations

In this chapter, an overview of the most relevant theoretical aspects that concern the analysis work is presented. The chapter is structured as follows. Firstly, the basis of Quantum Chromodynamics, the fundamental theory of strong interaction, are discussed in Sec. 2.1. Secondly, an overview of the physics of ultra-relativistic hadronic collision, such as those happening at the LHC, is presented in Sec. 2.2. Thirdly, the use of Monte-Carlo generators to describe particle production in hadronic collisions is addressed in Sec. 2.3. Fourthly, an introduction to several phenomenological approaches to describe inclusive charged particle production in proton-ion collisions are discussed in Sec. 2.4. Finally, the goal of the data analysis is introduced in Sec. 2.5, along with the observables to be measured and some essential definitions.

2.1 The Theory of Quantum Chromodynamics

The Standard Model (SM) is a theory of fundamental microscopic interactions that successfully explains most of the known phenomena in elementary particle physics [14–17]. The theory is a consistent framework that allows physical observables to be computed up to a precision limited by the current technical capabilities. The SM explains all so far observed microscopic phenomena, which can be attributed to strong, electromagnetic or weak interactions. The SM is structured as a renormalisable field theory based on a local symmetry around 14 conserved currents. These correspond to 8 colour charges, related with the strong force, and 4 electro-weak charges (including the electric charge) that are

related with the weak and electromagnetic interactions. This structure gives rise to the $\mathrm{SU}(3) \otimes \mathrm{SU}(2) \otimes \mathrm{U}(1)$ algebra.

In the SM the matter fields, all of spin 1/2, come in three generations or families with identical quantum numbers but different masses. On the one hand, a pair of quarks per family: (u, d), (c, s), (t, b). Quarks constitute protons, neutrons and all hadrons, and carry colour and electro-weak charges, thus experimenting the three mentioned interactions. On the other hand, leptons carry electro-weak charges but no colour, thus are invisible to the strong force. These are the electron e^-, the muon μ^-, the tauon τ^- plus the three associated neutrinos ν_e, ν_μ and ν_τ. All particles have its corresponding antiparticle with the same mass but all charges with opposite sign.

The SM is completed by 14 force carriers (one per conserved current) of spin 1. Eight of them are gluons g that mediate the strong interactions and carry themselves colour charge. From the electro-weak sector, these are the photon γ and the weak interaction bosons W^+, W^- and Z. Force carries are massless, except for the massive W^+, W^- and Z bosons with $m_{W^\pm} = 80.379 \pm 12\,\mathrm{GeV}/c^2$ and $m_Z = 91.1876 \pm 21\,\mathrm{GeV}/c^2$ [16]. This feature explains why the weak force is short-range, contrary to the electromagnetic interaction. The mass of these bosons also indicates that the electroweak gauge symmetry has been broken. In the SM, there is an additional spin 0 particle in charge of spontaneously breaking this symmetry, the Higgs boson H^0.

Quantum Chromodynamics (QCD) [1] corresponds to the SU(3) component of the SM and is the theory that describes the strong interactions of the coloured quarks and gluons. The properties of the strong interaction are of particular interest to contextualise the present work, since this force is behind the behaviour of hadrons. The lagrangian of QCD is given by

$$\mathcal{L} = -\frac{1}{4}\sum_{A=a}^{8} F^{A\mu\nu}F^{A}_{\mu\nu} + \sum_{j=1}^{n_f} \bar{q}_j(i\not{D} - m_j)q_j. \tag{2.1}$$

Here, q_j are the quark fields (of n_f different flavours) with mass m_j; $\not{D} = D_\mu\gamma^\mu$, where γ^μ are the Dirac γ-matrices and D_μ is the covariant derivative:

$$D_\mu = \partial_\mu - ie_s\sum_A t^A g^A_\mu; \tag{2.2}$$

and e_s is the gauge coupling; g^A_μ correspond to the gluon fields, with A running from 1 to $N_c^2 - 1 = 8$, *i.e.* eight kinds of gluons and t^A matrices correspond to eight 3×3 matrices that are the generators of the SU(3) colour group. These generators fulfil $[t^A, t^B] = iC_{ABC}t^C$, where C_{ABC} are the complete asymmetric structure constant of SU(3). In the first term of Eq. 2.1, $F^A_{\mu\nu}$ is the field tensor, which describes the dynamics of the gluon field, and is given by

$$F^A_{\mu\nu} = \partial_\mu g^A_\nu - \partial_\nu g^A_\mu - e_s C_{ABC} g^B_\mu g^C_\nu. \tag{2.3}$$

The terms of the QCD lagrangian of Eq. 2.1 describe the possible types of couplings among the coloured particles. The physical vertices in QCD include: gluon-quark-antiquark, 3-gluon and 4-gluon vertices. The 3-gluon and 4-gluon vertices are not present

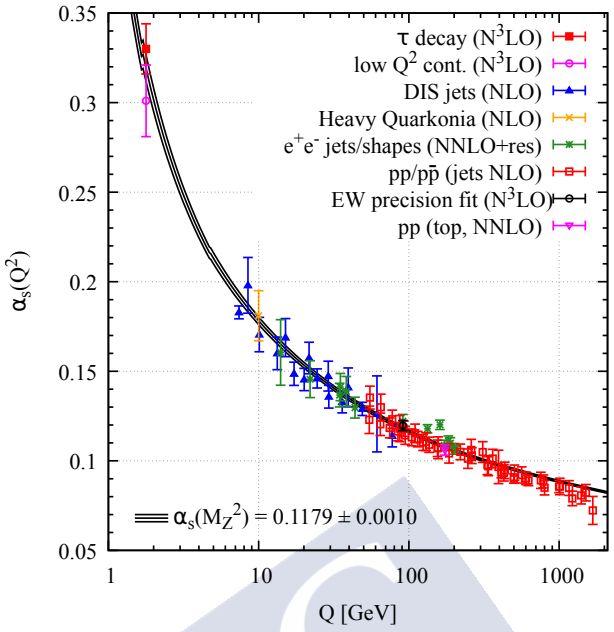


Figure 2.1: Dependence of α_s with the energy scale Q from different measurements [16]. In brackets, the degree of perturbation theory used to extract α_s is indicated (NLO: next-to-leading order; NNLO: next-to-next-to-leading order; NNLO + res.: NNLO matched to a resummed calculation; N^{3}LO: next-to-NNLO).

in other sectors of the SM like for photons (Quantum Electrodynamics). This property explains many of the differences between the strong and electromagnetic interactions.

The quantity e_s (or $\alpha_s = e_s^2/(4\pi)$) is the QCD coupling constant, and is the only fundamental parameter in QCD besides quarks masses. Although termed "QCD coupling constant", α_s is not constant for every vertex but depends on the transferred four momentum squared, Q^2, between the interacting particles. Fig. 2.1 shows the dependence of α_s with Q as measured in different experiments in the range $1 < Q < 10^3$ GeV/c as of 2020 [16]. Examples of these complementary measurements are hadronic decays of the τ^-, the observed spectra of bound states of heavy quarks ($b\bar{b}$ or $c\bar{c}$ or quarkonia) or jet production in e^-e^+.

The behaviour of α_s determines two of the core phenomena of QCD: asymptotic freedom and confinement. For large Q, α_s decreases and as consequence the strong interaction becomes weaker, so for $Q \to \infty$ quarks and gluons experiment no strong interaction. This is a common property of all gauge theories based on a non-commuting group of symmetry, and is known as asymptotic freedom. Since the coupling decreases asymptotically, QCD related quantities can be computed using perturbation theory in this regime. This approach will be introduced in Sec. 2.1.1.

The opposite behaviour is observed for low Q, where α_s grows rapidly. For values of order $Q \lesssim \Lambda_{\text{QCD}}$, where Λ_{QCD} is an energy of order a few hundred MeV, the interaction

becomes large and perturbation theory cannot be applied. In this regime the QCD interaction potential between colour charges increases linearly with the distance between the charges. This property can be studied in the lattice, which is a reformulation of QCD on a discrete space time, allowing this non-perturbative behaviour to be addressed [18]. The main consequence is that particles with colour charges will appear in the physical spectrum as composite objects which are colour-neutral, which are called hadrons. The most basic hadron type is a meson, composed of a colour neutral pair $q\overline{q}$. If one tries to separate the quark and the antiquark the interaction energy grows until new $q\overline{q}$ pairs emerge from the vacuum and new neutral mesons are coalesced instead of free quarks.

More complex examples of hadrons are baryons, which are qqq or $\overline{qqq}$ states with compensating colour charges. Mesons and baryons constitute the bulk of the known hadron spectra, for instance protons and neutrons are baryons composed of uud and ddu quarks respectively. Hadron structure is highly complex: along with the mentioned quarks, which are called valence quarks, there are a multitude of quark-antiquark pairs and gluons constantly interacting among themselves. They are very important to the point that most of the hadron mass originates from their contribution. A consequence of this complex structure is that when an elementary particle interacts inelastically with a hadron, it is actually interacting with one of its elementary components. In this context they are called partons in general, since they carry *part* of the total hadron momentum.

2.1.1 Perturbative QCD

One of the main applications of QCD is the computation of different particle production cross-sections in particle colliders. The approach of perturbative QCD (pQCD) [2,3] consists in expanding the cross-section in powers of the constant α_s. This approach is valid for hard processes with $Q^2 \gg {\Lambda_{\mathrm{QCD}}}^2$ so that $\alpha_s \ll 1$.

In the context of this thesis the most interesting processes are hadronic collisions, and in particular proton-proton interactions are one of the simplest cases. However, these are already challenging due to the composite nature of the proton that derives from the property of confinement as was explained above. Intrinsically, the proton structure involves non-perturbative processes that cannot be addressed with pQCD. However, the collinear factorisation theorem [2] can be used to compute production cross-sections of particles produced in hard processes where $Q^2 \gg {\Lambda_{\mathrm{QCD}}}^2$. The approach allows to express the inclusive production cross-section of a hard elementary particle k in a collision of hadrons h_1 and h_2 as

$$\mathrm{d}\sigma^{h_1+h_2\to k+X}(\mu^2, Q^2) = \sum_{i,j,X'} f_{i/h_1}(x_1, Q^2) \otimes f_{i/h_2}(x_2, Q^2) \otimes \mathrm{d}\hat{\sigma}^{ij\to k+X'}(\mu^2, Q^2). \qquad (2.4)$$

Here, $f_{i/h_1}(x_1, Q^2)$ and $f_{i/h_2}(x_2, Q^2)$ are the parton distribution functions (PDFs), which describe the number density distribution of partons i (j) in a hadron h_1 (h_2) at a momentum fraction x_1 (x_2) and a factorisation scale Q^2. The momentum fraction x of a parton within a proton is defined as the fraction of the proton momentum carried by the

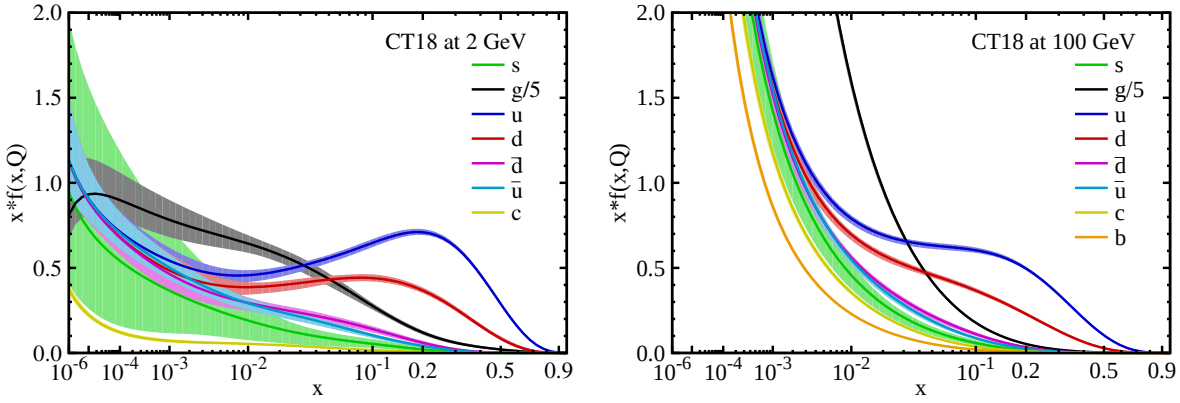


Figure 2.2: Parton distribution functions from the CT18 analysis [23] at $Q = 2\,\text{GeV}$ (left) and $Q = 100\,\text{GeV}$ (right), considering natural units, for u, $\overline{u}$, d, $\overline{d}$, $s = \overline{s}$, $b = \overline{b}$ and g. The gluon PDF is scaled as $g(x, Q)/5$ in all instances, and the charm distribution $c(x, Q)$ is perturbatively generated by evolving from $Q_0 = 1.3$ and $1.4\,\text{GeV}$. The band around each contribution represents the estimated uncertainty.

parton. The quantity $\mathrm{d}\hat{\sigma}^{ij\to k+X'}(\mu^2, Q^2)$ is the production cross-section from the process where partons i and j from h_1 and h_2 interact and yield the elementary particle k and any additional products X', considering the interaction diagrams up to the targeted order of the prediction. Note that the equation is summed over i, j and X', and therefore all combinations of partons and possible products have to be considered. The dependence with μ is related with the normalisation scale. The differential cross-section of Eq. 2.4 needs to be integrated over the values of x_1 and x_2 that contribute to $\mathrm{d}\hat{\sigma}^{ij\to k+X'}(\mu^2, Q^2)$ for a given final state with particular kinematics.

The PDFs describe the parton composition of the proton depending on (x, Q^2). They are universal non-perturbative objects with scale Q^2 evolution driven perturbatively by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [19–21]. Currently they cannot be computed from first principles in QCD, and they are determined from global analyses of data for deep inelastic lepton-nucleon scattering and from related hard-scattering processes initiated by nucleons. A recent review of the current status of PDFs can be found in Ref. [22]. As example, Fig. 2.2 shows the parton distribution functions from the CT18 analysis [23] at $Q = 2\,\text{GeV}$ and $Q = 100\,\text{GeV}$.

The case of study for this thesis is inclusive high-p_T hadron production. The expression in Eq. 2.4 is the cross-section for an elementary particle, but hadrons are composite states so the equation is not directly applicable for hadron production. However, new partons can be produced in the hard scattering and then hadronise to colourless hadrons that can be detected. The probability for a produced parton to hadronise in a hadron is encapsulated in the parton-to-hadron fragmentation functions (FFs) [24]. Therefore, the final inclusive hadron cross-section is obtained by convolution of the hard parton spectra

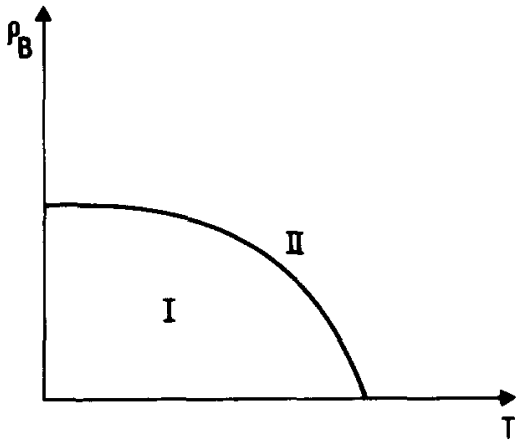


Figure 2.3: QCD phase diagram. Figure taken from Ref. [31].

from Eq. 2.4 with the FFs:

$$\mathrm{d}\sigma^{h_1+h_2\to h+X}(\mu^2, Q^2, Q_F^2) = \sum_k \mathrm{d}\sigma^{h_1+h_2\to k+X}(\mu^2, Q^2, Q_F^2) \otimes D_{h/k}(z, Q_F^2). \tag{2.5}$$

Here, z describes the momentum fraction carried away by the hadron h from the parent parton k; $D_{h/k}(z, Q_F^2)$ is the parton-to-hadron fragmentation function; and Q_F^2 is the fragmentation scale. As with Eq. 2.4, to obtain inclusive cross-sections for particles with specific kinematics the integration over the contributing x_1 and x_2 must be performed. As with PDFs, FFs are generally obtained through global analyses of experimental data. A recent review on FFs can be found in Ref. [25]. There are many examples of analyses of fragmentation functions for different hadron species [26–28].

2.1.2 The QCD phase diagram

The property of asymptotic freedom explained before evidences the existence of a new high temperature phase of weakly interacting quarks and gluons, termed quark-gluon plasma (QCD) (see Ref. [29, 30] and references therein). This state of hadronic matter would occur when the temperature is high enough that thermal interactions between quarks and gluons have $Q^2 \gg {\Lambda_{\mathrm{QCD}}}^2$ so they would asymptotically free.

A sketch of the present-day knowledge of the QCD phase diagram is shown in Fig. 2.3. The different phases of strongly interacting matter are presented as a function of the temperature T and the baryon chemical potential μ, which represents the net number of baryons with respect to anti-baryons.

At zero temperature and chemical potential the interaction between quarks is dominated by large distances and the coupling α_s is large. As a result, quarks and gluons

are confined in hadrons, with a mass close to $\Lambda_{\mathrm{QCD}} \sim 200\,\mathrm{MeV}$. For instance, the proton has a mass of $m_p \approx 935\,\mathrm{MeV}$. At very high temperature ($T \gg 170\,\mathrm{MeV}$), quarks and gluons have thermal momentum $p \sim T \gg \Lambda_{\mathrm{QCD}}$, and therefore they will interact weakly among themselves forming a plasma-like state of colour charges, called Quark Gluon Plasma (QGP). Since colour confinement is not present in the QGP, a phase transition must be separating this deconfined state from the hadronic gas. Studies in Lattice QCD have shown that this transition, at least at low values of μ, is a crossover [32], which corresponds to a rapid change instead of a jump.

The QGP is assumed to be formed naturally in the early universe, when the universe was a few microseconds old and the temperature was above Λ_{QCD}, too hot for hadron formation. It is also possible to study the QGP and the phase transition using ultra-relativistic collisions of heavy-ions [9]. Collisions of AuAu at the Relativistic Heavy Ion Collider (RHIC) have reached a nucleon-nucleon centre-of-mass energy of $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{TeV}$ in the early 2000, while at Large Hadron Collider (LHC) PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ were performed in the early 2010s. In Sec. 2.2.2, a more detailed description of the observations that led to the characterisation of the state formed in such collisions is given.

2.2 Ultra-relativistic hadronic collisions

Particle collisions are a fairly intuitive approach to investigate the structure of matter and the fundamental interactions among its constituents. As was described in Sec. 2.1, hadrons are composite objects of partons with a high degree of complexity. Here, the discussion is focused in high-energy collisions with energies in the centre-of-mass system $\sqrt{s} \sim 1\,\mathrm{TeV}$. This is the energy range of the LHC.

In Sec. 2.1, it was seen how the behaviour of QCD changes dramatically with the scale of the interaction given by Q^2. In this context, it is interesting to distinguish two types of processes depending on the applicability of pQCD in the involved partonic interactions. In hard processes there is a large momentum transfer and the interaction is within the hard scale $Q^2 \gg \Lambda_{\mathrm{QCD}}$ or $Q^2 \gtrsim 1\,\mathrm{GeV}/c^2$. As explained in Sec. 2.1.1, they can be described using pQCD. Due to the large Q^2, there is a large available amount of energy in the final state allowing heavy particles to be produced. These processes are the source of most physics analyses at LHC that search for New Physics beyond the Standard Model (NP). Such searches can be direct, which are searches of new fundamental heavy particles, or indirect, which attempt to infer the presence of NP in lower Q^2 processes by measuring discrepancies in observables with the SM prediction.

All the processes that do not reach the hard scale are classified as soft. They are non-perturbative, and they are much less understood than hard processes. However, they constitute a large fraction of what happens in a hadron-hadron collision and have to be modelled in most physics analyses since they may appear as backgrounds. Also, hard and soft processes are superimposed in the same hadron-hadron event, which makes the

separation between signal and background even harder.

A more detailed study of hadron-hadron collisions will be done in the following sections. Most details will be introduced first for the case of proton-proton collisions (Sec. 2.2.1) since they are the simplest case that will be covered in this thesis. Next, an introduction to the more complex heavy-ion collisions will be made in Sec. 2.2.2, focusing onthe topic of proton-ion collisions.

2.2.1 Proton-proton collisions

It is interesting to do a more detailed classification of the types of processes between two protons. Such interactions can be classified by the characteristics of the particles in the final-state of the process. Before defining these categories, the concept of rapidity of a particle must be introduced, defined as

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}. \tag{2.6}$$

Here, E is the energy of the particle and p_z is the momentum component of the particle parallel to the trajectory of the interacting protons. A particle with $y = 0$ has a transverse trajectory to the collision axis and is said to be in the central rapidity region, while a particle with $y \gg 0$ ($y \ll 0$) is said to be in the forward (backward) rapidity region. In the limit $y \to \infty$ the particle is parallel to the beam direction, being the sign of y chosen by convention with respect to one of the incident protons. The variable is very appropriate to describe the kinematics of the produced particles in the collision. A disadvantage of this quantity is that in order to calculate the rapidity, the mass of the final-state particle must be known to infer its energy. For this reason, the alternative variable called pseudorapidity,

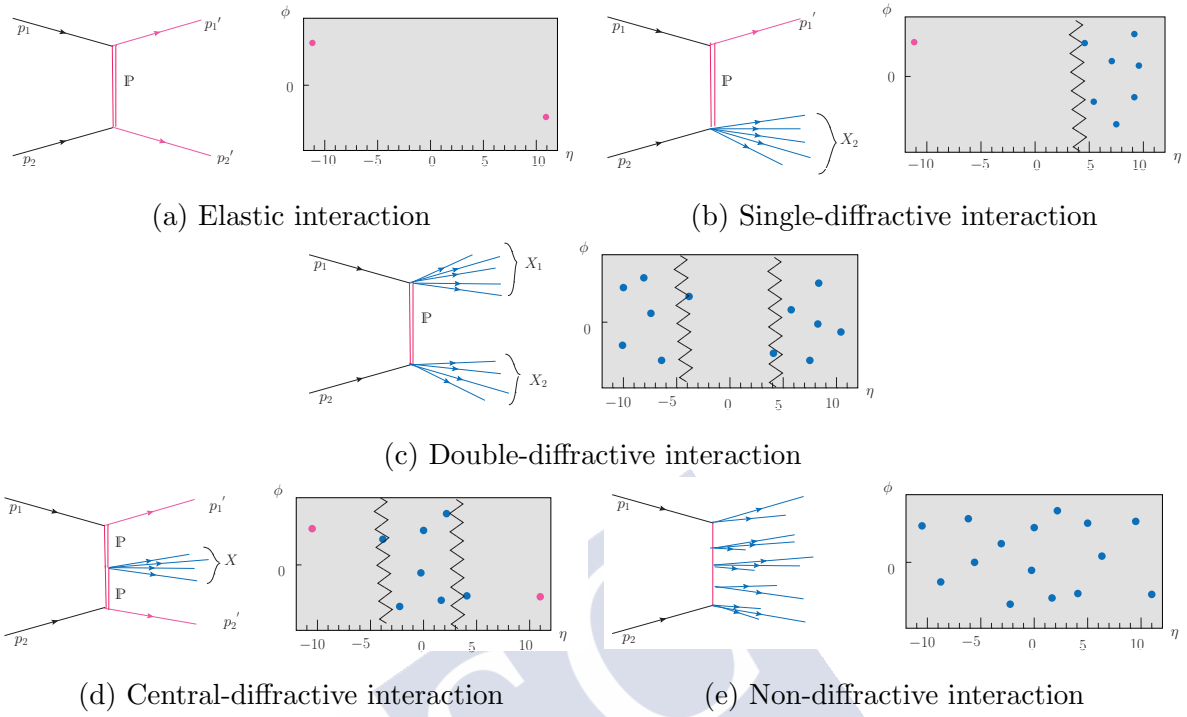
$$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \tag{2.7}$$

where θ is the angle between the particle and the trajectory of the colliding protons, is used. For massless particles, both quantities are equal and for $p_{\mathrm{T}} \gg m$ particles, $\eta \approx y$.

The first possibility when two, p_1 and p_2, protons collide is that they interact and emerge intact without any additional particle in the final state: $p_1 + p_2 \to p_1 + p_2$. These are called *elastic* interactions. Fig. 2.4a shows a diagram of the interaction, which takes place with a Pomeron exchange, which is a colourless QCD state designed with $\mathbb{P}$. The particle distribution in the (η, ϕ) plane only shows two protons slightly deviated from their initial trajectory so they appear at very low θ. At 7 TeV, the elastic proton-proton cross-section is 25.4 ± 1.1 mb, about a fourth of the total cross-section of 98.6 ± 2.2 mb [34].

When additional particles appear in the final state, the interaction is called *inelastic*. At 7 TeV, the inelastic proton-proton cross-section is 72.9 ± 1.5 mb [35]. Inelastic processes are usually classified in diffractive and non-diffractive (ND).

In diffractive processes [36] one of the incoming protons (or both) dissociates into two or more final state particles with mass much lower than $\sqrt{s}$ and the proton quantum



(a) Elastic interaction

(b) Single-diffractive interaction

(c) Double-diffractive interaction

(d) Central-diffractive interaction

(e) Non-diffractive interaction

Figure 2.4: Possible types of interactions in proton-proton collisions. The diagrams in the right indicate the distribution of particles in the final state in the (η, ϕ) plane to represent the presence of LRG. Figures taken from Ref. [33].

number. Experimentally they are linked to the presence of Large Rapidity Gaps (LRG) in the distribution of particles in the final-state, which are regions of η without any particle. Their size is not well defined, but typically $\Delta\eta > 4$. They usually involve very low Q^2 and therefore they are non-perturbative, although hard diffractive processes are also possible.

Several types of diffractive processes are possible. In a single-diffractive (SD) interaction, shown in Fig. 2.4b, one proton dissociates and the other stays intact: $p_1 + p_2 \to X + p_2$, where the + symbol represents a LRG. In double-diffractive (DD) interaction, presented in Fig. 2.4c, both protons are dissociated: $p_1 + p_2 \to X + Y$. Finally, in a central diffractive (CD) process (Fig. 2.4d), the protons stay intact but an object X is produced in the central region: $p_1 + p_2 \to p_1 + X + p_2$. The object X must be colour-neutral object, for instance a J/ψ meson.

In non-diffractive (ND) processes, both protons interact and break-up, producing a multitude of particles in the final state that span over the entire available rapidity range. These are presented in Fig. 2.4e. The total proton-proton interaction cross-section can be calculated by adding the interaction cross-sections for all the possible processes:

$$\sigma_{\text{total}} = \sigma_{\text{elastic}} + \sigma_{\text{inel}} = \sigma_{\text{elastic}} + \sigma_{\text{diff}} + \sigma_{\text{ND}} = \sigma_{\text{elastic}} + \sigma_{\text{SD}} + \sigma_{\text{DD}} + \sigma_{\text{CD}} + \sigma_{\text{ND}}. \quad (2.8)$$

The contribution of diffractive processes is large, up to 40%. However, most particle production in minimum bias proton-proton collision, this is, without any preference

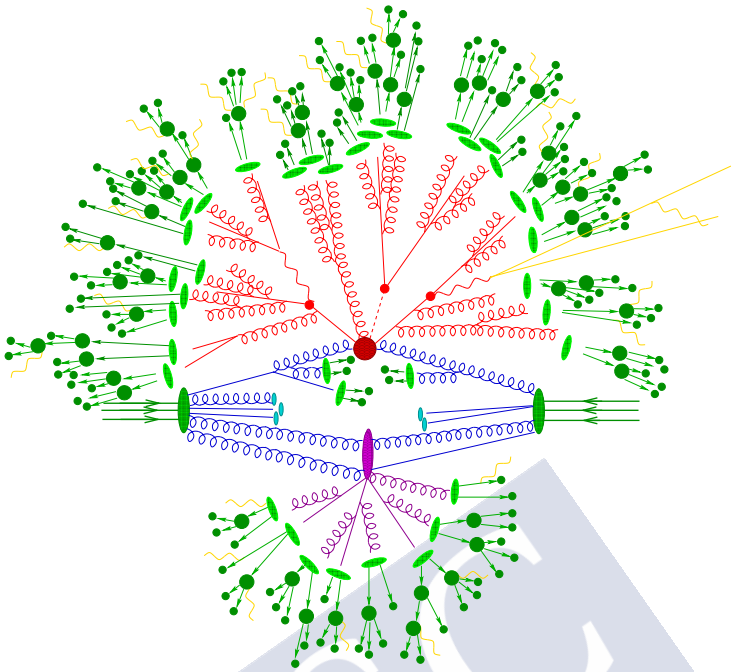


Figure 2.5: Representation of a proton-proton interaction with a hard scattering with $t\bar{t}H^0$ production (big red circle), and a additional parton interaction (purple blob). Figure taken from Ref. [37].

for any event geometry, comes from non-diffractive interactions. This is even truer for particles of high p_{T}, since diffractive processes usually involve interactions with lower Q^2. Nevertheless, the understanding of diffractive processes is important to distinguish both contributions. This is specially true in high pile-up conditions, this is, when more than one proton-proton interaction occurs in a given event and diffractive and non-diffractive signatures overlap in the final state. Due to the difficulty to separate diffractive and non-diffractive events, generic inelastic interactions are considered in the rest of the thesis.

In general, in a pp collision hard and soft processes are superimposed. To illustrate this, a generic proton-proton interaction with a hard parton scattering is presented in Fig. 2.5. In the hard interaction, represented with a big red circle, a $t\bar{t}H^0$ event is produced. The lines of the three particles are followed by their respective decay. Additional gluon radiation is emitted and is drawn in red. As gluons carry themselves colour charge, they might radiate other gluons leading to showers of partons. These partons will undergo a hadronisation process indicated in light green blops in the figure, forming colour neutral hadrons in the final state. Some of the formed hadrons (big green circles) will decay into more stable hadrons (small green circles). The whole process is initiated usually from a high momentum parton produced in the hard scattering, and the shower of hadrons in a definite (η, ϕ) region is called a *jet*. The hadronisation process involves low Q^2, and is generally addressed with phenomenological models.

Gluon and photon radiation can be emitted by any produced particle with colour or electric charge respectively. Photon radiation is drawn in yellow in the figure. The radiation is classified in initial-state radiation (ISR) or in final-state radiation (FSR), depending the stage of the collision when it is emitted.

On top of the primary hard interaction, it is possible that additional pairs of partons of the colliding protons experiment other hard (or semi-hard) interactions. These processes are called in general Multiple Parton Interactions (MPI). Events in which two pairs of partons interact are called Double Parton Scattering (DPS). In Fig. 2.5 an additional secondary partonic interaction occurs and such process is drawn as a purple blob. before the final-state partons hadronise (light green blobs) and hadrons decay (dark green blobs).

All the particles produced jointly with the hard process are usually called *underlying event*. The underlying event includes mostly particles from ISR, FSR and DPS. Another phenomenon to be mentioned is parton and hadron rescattering in the final state, usually referred to as final-state interactions. These are specially important at high particle densities, therefore in heavy-ion collision, where the mean particle multiplicity is larger.

It can be concluded that in *pp* collisions there is a multitude of processes, soft and hard, that influence the particle distribution in the final state. Even if one is interested in studying hard interactions, which are generally the target of New Physics searches, these will always appear overlaid with a multitude of particles from other interactions. There is not yet a fully predictive framework to address those, and they are usually modelled in Monte-Carlo generators. A description of this approach is made in Sec. 2.3. The complications only grow when moving to heavy ion collision, since the number of interacting partons scales and the complexity of the problem does as well. More information about them is given in Sec. 2.2.2.

2.2.2 Heavy-ion collisions

In collisions of two heavy-ions (or AA collisions), the proton-proton picture described above becomes more complex [9]. A ultra-relativistic heavy nucleus such as lead Pb is a Lorentz contracted disc of thickness about $14/\gamma\,\mathrm{fm}$ with $\gamma \approx 2500$ at LHC energies. The disk is formed by a highly complex system of partons with a momentum distribution close to a superposition of the individual nucleons but with small modifications due to the interactions with the rest of the nucleons. The deviations with respect to the incoherent superposition of nucleons configure what is called the initial stage of the collision.

When the two discs overlap, many soft interactions with small Q^2 between the partons are produced, and also rarer hard interactions which lead to particle production with high transverse momentum. After the collision, the energy density is extremely high, with temperatures above the crossover temperature described before in Sec. 2.1.2. In this situation, quarks and gluons cannot be described as a collection of distinct individual hadrons, but more as a QGP. This state is shown in red in Fig. 2.6 at $5\,\mathrm{fm}/c$ after the collision. However, these quarks and gluons are still strongly coupled and the system behaves as a relativistic hydrodynamic fluid with a very low viscosity to entropy density

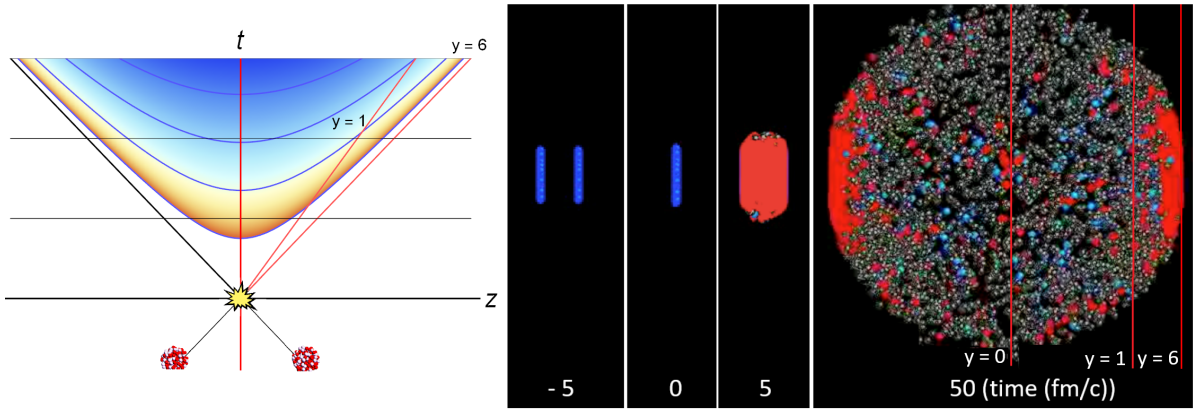


Figure 2.6: (Left) Representation of a heavy ion collision showing the time evolution and the expansion of the system in the coordinate parallel to the trajectory of the ions in the vertical and horizontal axes respectively. Colour gives and indication of the temperature of the system. Blue curves indicate constant proper time. (Right) Snapshots of a central 2.76 TeV PbPb collision at different times. Hadrons correspond to blue and gray spheres, while QGP is coloured in red. In both figures, the hottest regions are found at high rapidity, close to the outgoing remnants of the nuclei. Figure taken from Ref. [9].

ratio. With time, both disks recede from each other and the QGP at central rapidities expands driven by strong hydrodynamic pressure and cools down, eventually hadronising into baryons and mesons (blue and grey spheres in the figure). The region at forward and backward rapidities, closer to the nuclei remnants, is hotter and continues forming QGP for a longer time as shown in Fig. 2.6.

The previously described evolution of nucleus-nucleus collisions was constructed historically with a series of historical experimental observations. At CERN, using Pb beams, a strong J/ψ suppression with respect to proton-proton [38] was observed, and also an enhanced production of strange hadrons [39]. Then, at RHIC since the early 2000s, collisions of heavy ions up to $\sqrt{s_{\rm NN}} = 200\,\text{GeV}$ were achieved. Here, a strong azimutal asymmetry (elliptic flow v_2) in particle distribution was observed [40], which agreed with the predictions from hydrodynamics, introducing the idea of a strongly interacting plasma.

Another important observation was the suppression of hadron production at high $p_{\rm T}$ with respect to proton-proton [10]. The strong hadron suppression is usually understood as a result of jet quenching [11], which is a modification of the jet properties due to medium-induced radiation. Predictions of the production of high-$p_{\rm T}$ jets in the primary hard parton interaction are feasible in the pQCD framework (see Sec. 2.1.1), and with comparisons with pp data the properties of the medium that induces the modification can be inferred. The same observations were reproduced at the LHC, by ALICE [41], CMS [42] and ATLAS [43].

To evaluate reliably all these observations, a critical point is to provide a baseline to ensure that the effect was originated in the dense medium and not as a initial-stage

modification of the partonic distributions in the nuclei. Here, proton-nucleus (pA) collisions provide an ideal testing ground, since the energy density is lower than that in AA collisions, and in principle not enough for QGP formation. Therefore, effects derived from interactions with the hot medium should not be present and the modification with respect to pp can be linked other "cold" phenomena. Those are usually encompassed in the term Cold Nuclear Matter (CNM) effects. They will be discussed in Sec. 2.4 due to their importance for the object of this thesis.

Nevertheless, recent experimental observations [44,45] have shown that the differences between AA, pA and even pp collisions are more diffuse, complicating the simple picture of hot and cold effects described above. A striking observation is that signs of collectivity, which were assume to be intrinsic to the QGP picture, have been found in small systems like pp and pA [46]. Some models link these findings to modifications in the initial stage [47], but opening the door for these modifications to be present also in AA collisions. In such case, they would need to be accounted in the description of the QGP in AA collisions.

2.3 Monte-Carlo generators

Although QCD provides a complete description of the strong interaction, the non-perturbative behaviour towards low Q^2 reduces its predictive power, as explained in Sec. 2.1. Unfortunately, as was described in Sec. 2.2, many processes in hadronic collisions are soft and therefore non-perturbative. Even for the case of hard interactions, they are inherently linked to other non-perturbative processes, for instance hadronisation. In spite of these difficulties, the access to accurate descriptions of the final state of hadron collisions is of utmost importance to every study carried out at LHC. The solution is to use Monte-Carlo (MC) generators, which are models that integrate the probability of the different phenomena to reproduce the final-state of hadronic collisions [8]. Defining the probabilities of processes to occur is the basis of the construction of a MC generator. Hard processes can usually be computed accurately from first principles and therefore their probabilities are accessible, but soft processes need to be modelled.

The use of MC generators is crucial for the development of the physics program in hadron colliders like the LHC [8]. Monte-Carlo predictions are used to simulate the signature of a process of interest or *signal*, which can be a known SM process or a New Physics signature, and the expected signal-like signatures originated in other SM processes that can disturb the observation or *background*. MC predictions are fundamental not only to conduct physics analyses, but to study the sensitivity of particular studies for the planification of new accelerators. In detector development, a precise simulation of the expected final-state particles is crucial to define in advance the technical specifications needed to perform a particular physics measurement.

In addition to being able to model correctly signal and background channels, the detector response has to be reproduced to correct for effects that disrupt the target ob-

servables such as detection efficiencies or detector-induced backgrounds. For this purpose, software as GEANT4 [48], specialised in propagating the final-state particles of the hadronic collision into the signatures in the detection systems, is used. Such propagation is performed according to the well-known interactions between high-energy particles and matter.

The importance of MC generators goes beyond collider physics. For instance, they are widely used in cosmic ray physics to model the interaction of high energy cosmic rays with the atmosphere producing extensive air showers [49]. These showers are detected by experiments such as the Pierre Auger Observatory [50], and are the only way to measure cosmic rays of energy larger than 10^{15} eV. Currently, the dominant source of systematic uncertainty in the interpretation of these measurements is linked to how reliable the simulations are, specially those for hadronic interactions [5, 6]. In addition, there are inconsistencies between measurements and simulations in the number of muons produced in such cascades, and they could be due to a poor simulation of hadronic observables at LHC energies, specially in forward rapidities [7].

A very important part of developing accurate MC generators is called *tuning*. As was mentioned, a rather large fraction of the processes occurring in the event cannot be computed from first principles and has to be modelled with phenomenological considerations. In general, such models will have a number of parameters to be adjusted by comparisons with experimental data. This validation of the models must be global to ensure that the underlying physics mechanisms are correctly described, as the model should be more than just a parametrisation of data. Since generators include many different processes, the number of parameters is usually of the order of 15 or more, so they are grouped depending on the processes involved and are tuned in different stages. The order is usually the following: first, hadronisation and final-state fragmentation processes; second, initial-state parton showers; and third, multiple parton scattering and beam remnant effects. Packages like Rivet [51] allow for a systematic tuning of such parameters according to extensive libraries of experimental data.

Nowadays, there are many MC generators able to reproduce pp, pA and AA collisions. In the following, the PYTHIA generator for pp and the EPOS generator for heavy-ion collisions will be described. These two models are integrated in the LHCb simulation framework and are used in the analysis of the present thesis (see Sec. 3.2.5).

PYTHIA. A very standard general-purpose event generator for proton-proton collisions is PYTHIA [52]. PYTHIA is one of the most used event generators at LHC. PYTHIA has implemented an extensive list of hard processes of interactions between partons, and the parton level cross-section is then convoluted with PDFs to estimate event rates, similarly to what was explained in Sec. 2.1.1. On top of the primary hard interaction, ISR, FSR and MPI, which were described in Sec. 2.2.1, are also modelled. Diffractive processes are handled in the Ingelman–Schlein picture [53], and are considered on top of non diffractive events to reproduce the inelastic pp cross-section. The ISR and FSR showers are ordered in descending transverse momentum [54], as well as MPI [55]. This means that they are

evolved perturbatively with a combined evolution equation. Hadronisation is based on the Lund string fragmentation framework [56]. PYTHIA has been tuned extensively to be used reliably by the LHC experiments. In the context of this work the LHCb tune [57] is used, tailored for the LHCb forward acceptance.

EPOS-LHC. Another event generator of pp, pA and AA collisions, is EPOS-LHC [58], which is specially suited to reproduce minimum-bias data. The generator is based on the EPOS 1.99 model [59], relying on the Parton-Based Gribov Regge Theory [60], where the interaction between the hadrons is seen as a exchange of a "parton ladder" [61]. A parton ladder is composed of a pQCD hard process plus initial and final state linear parton emission. In that model, collective hadronisation is implemented in a core-corona approach [62], in which for every event a part of the string segments from the initial conditions hadronises normally (corona) and a part is used to create a core with collective hadronisation. A treatment of collective flow effects is also included.

2.4 Phenomenology of proton-ion collisions

As introduced in Sec. 2.2.2, pA collisions at LHC act as a baseline for AA collisions, but they are also a door to new physics opportunities [63], such as the investigation of the heavy nucleus partonic structure. The first proton-lead run at LHC took place in September 2012 at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$, in preparation for a longer run in February 2013 at the same energy. A collection of phenomenological predictions targeting that run can be found in [12]. In 2016, another pA run took place at LHC with $\sqrt{s_{\mathrm{NN}}} = 8.16\,\mathrm{TeV}$, and phenomenological predictions were improved with respect to the first run [64]. In addition, experiments at RHIC have been studying pA and dA (deuterium-nucleus) collisions since the early 2000s.

As mentioned in Sec. 2.2.2, modifications of the particle production rate in pA collisions relative to binary scaled proton-proton collisions are usually referred to as Cold Nuclear Matter (CNM) effects. The following discussion of the phenomenology will be focused on inclusive prompt charged particle production, also referred to as inclusive charged hadron production, considering that hadrons (π, K, p) constitute the bulk of the final state particles in a hadronic collision. For this observable, CNM effects are generally linked to modifications of the initial state of the nucleus with respect to the proton, rather than to final-state effects which would take place in a later stage of the collision.

Since in principle CNM effects are originated in modifications of the parton distributions within the nuclei, they might variate with respect to (x, Q^2), similarly to the parton distribution functions of the proton (see Fig. 2.2). Therefore, it is important to investigate them all over the (x, Q^2) range. Also, since they are essentially non-perturbative phenomena, experimental input is paramount. For these reasons, the coverage of the different LHC and RHIC experiments in (x, Q^2) needs to be studied.

However, it is not straightforward to relate the final-state particle kinematics, usually

given in variables like (η, p_{T}), with the (x, Q^2) of the partons of the nuclei that experienced the hard scattering. An approach followed in Ref. [65] is to consider LO $2 \to 2$ kinematics. In this scope, a way to approximate x in the nuclei is

$$x \approx \frac{m_{\mathrm{T}}}{\sqrt{s_{\mathrm{NN}}}} e^{-\eta}, \tag{2.9}$$

where $m_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2}$ is the transverse mass of the produced hadron of mass m, pseudorapidity η and transverse momentum p_{T}. Eq. 2.9 should be understood as an indication of the minimum value of x probed in the nuclei, since the contribution to the hadron cross-section for a given (η, p_{T}) will include contributions over a range of x. Since a pA collision is asymmetric, it must be clarified that a positive η is taken with respect to the direction of the proton beam.[1] A more detailed discussion of the x coverage can be found in Refs. [66,67]. Nevertheless, Eq. 2.9 allows to state that lower values of x can be probed with higher $\sqrt{s_{\mathrm{NN}}}$ and in more forward η.

Fig. 2.7 shows the coverage of most heavy-ion experiments at RHIC and LHC that have measured inclusive charged particle production. The value of x is obtained with Eq. 2.9 and Q^2 is approximated as the transverse mass m_{T}. An average hadron mass of $m = 255\,\mathrm{MeV}/c^2$ has been considered, based on the proportion of π, K and p in the EPOS generator at low p_{T}.

In the figure, the yellow area labelled as "saturation region" correspond to the region where the gluon distribution in a heavy nuclei (Pb) could be saturated [72, 73]. This phenomena will be explained in Sec. 2.4.2. The dashed area, which indicates $p_{\mathrm{T}} > 1.5\,\mathrm{GeV}/c$, indicates the approximate frontier where the pQCD techniques start to become applicable. At LHC, the ALICE [41, 74], CMS [42, 68] and ATLAS [69] collaborations have measured prompt charged particle production at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ in the central region. At RHIC, measurements at more forward pseudorapidities have been performed by the BRAHMS [70] and PHENIX [71] collaborations with dAu collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. However, since for RHIC measurements $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$, the reached x is not as low as for LHC measurements.

The figure also shows the coverage of the LHCb experiment, which can probe $\eta > 0$ with pPb (the proton beam pointing towards the LHCb spectrometer) and $\eta < 0$ with Pbp (the lead beam pointing towards the LHCb spectrometer). More details about LHCb and its acceptance coverage will be given in Chap. 3.

In the following, a description of some theoretical approaches to describe charged particle production in pA collisions is presented. The objective is to present those that provide predictions that will be compared to the final results of this thesis. For more exhaustive lists, Refs. [12, 64] and references therein can be consulted.

[1] Note that the convention used by the ALICE collaboration is usually the oposite.

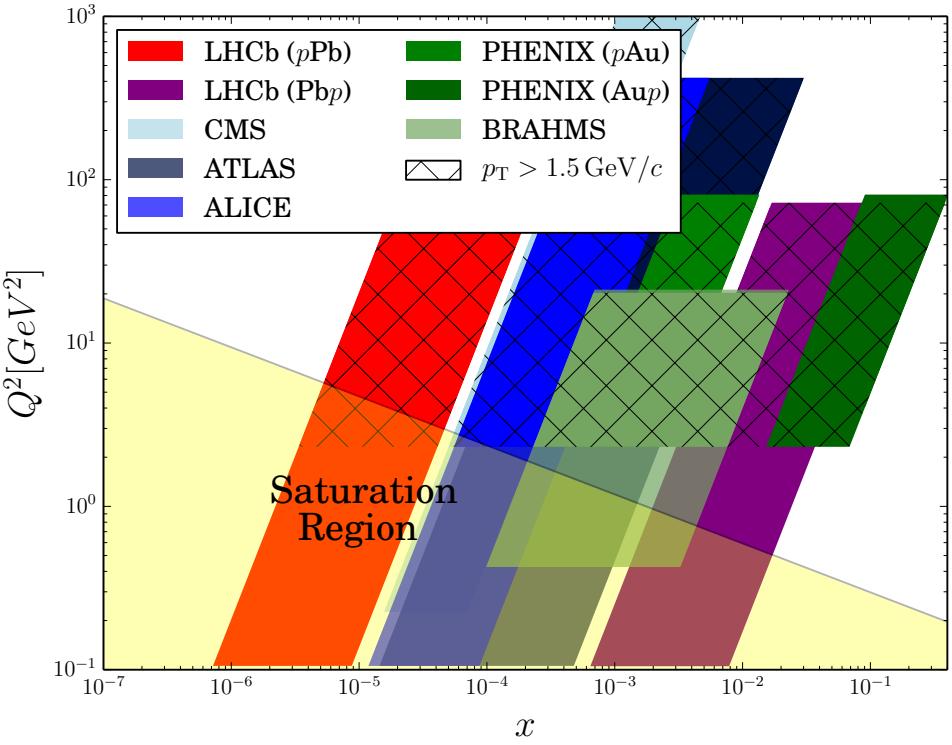


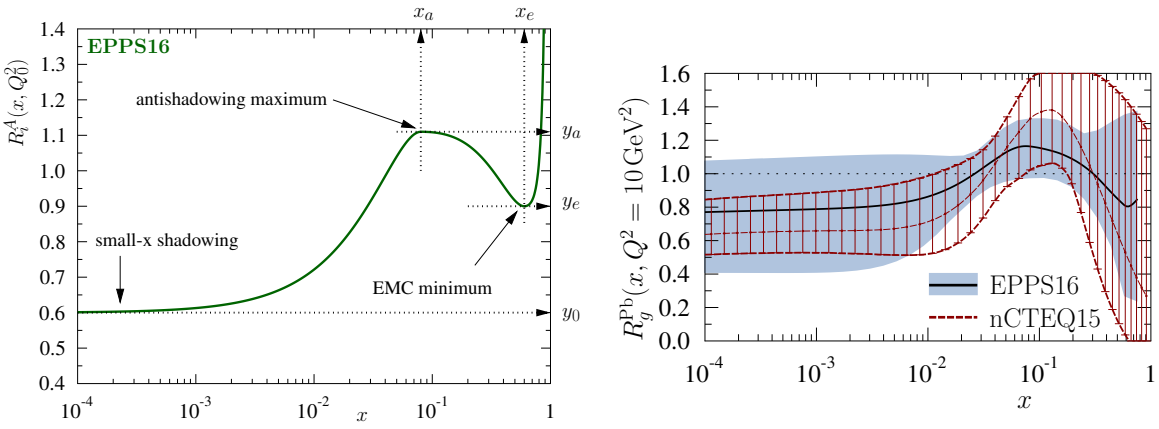
Figure 2.7: Kinematic coverage of LHCb and other experiments for pA or dA collisions in terms x from Eq. 2.9 and $Q^2 \approx m_{\mathrm{T}}^2$. The kinematic range is taken from measurements of inclusive charged particle production performed at the ALICE [41], CMS [68], ATLAS [69], BRAHMS [70], PHENIX [71] experiments.

2.4.1 Nuclear parton distribution functions

One way to characterise the partonic structure of the nuclei is to follow a similar approach as in Sec. 2.1.1, where the non-perturbative contribution to the scattering from the partonic structure of the proton was parametrised in the PDF $f_i^p(x, Q^2)$, where $i = q, \overline{q}, g$. In analogy, the PDF $f_i^A(x, Q^2)$ of a nuclei with A nucleons can be parametrised as:

$$f_i^A(x, Q^2) = A R_i^A(x, Q^2) f_i^p(x, Q^2) \tag{2.10}$$

where $R_i^A(x, Q^2)$ is the scale-dependent nuclear modification with respect to the proton PDF. Then, $R_i^A(x, Q^2)$ is parametrised at a value of $Q_0^2 \sim 1\,\mathrm{GeV}^2$ where pQCD can be applied reliably, and then the evolution for larger Q^2 is achieved with the DGLAP evolution equations. The parametrisation of $R_i^A(x, Q^2)$ has to include the nuclear size as an additional variable, not present in the PDF case, and is usually constrained with fits to the available experimental data. Then, $f_i^A(x, Q^2)$, which is called nuclear parton



(a) (b)

Figure 2.8: Fig. 2.8a: Fit function of nPDF set EPPS16 [75]. Fig. 2.8b: Gluon parametrisation for the lead nuclei at $Q^2 = 10\,\mathrm{GeV}/c^2$ for EPPS16 and nCTEQ15 [76] sets. Figures taken from Ref. [75].

distribution function (nPDF), can be used as an equivalent to a PDF in pQCD calculations to predict particle cross-sections in nuclear collisions using Eq. 2.4 and Eq. 2.5.

Nowadays, there are available many nPDF parametrisations [75–77], which differ on several details. For instance, the form of the parametrisation at the initial scale Q_0^2, the use of different sets of experimental data, the order of DGLAP evolution (LO, NLO...), the proton PDF used as reference, how isospin effects are treated (difference between neutron and proton), *etc.* As an example of the form of the initial parametrisation, Fig. 2.8a shows the fit function used in the EPPS16 analysis [75]. In this function, four main regions of the behaviour of $R_i^A(x, Q^2)$ can be distinguished [65], which correspond to:

- $R_i^A(x, Q^2) > 1$ for $x \gtrsim 0.8$: the Fermi motion region.
- $R_i^A(x, Q^2) < 1$ for $0.3 \lesssim x \lesssim 0.8$: the EMC region.
- $R_i^A(x, Q^2) > 1$ for $0.1 \lesssim x \lesssim 0.3$: the antishadowing region.
- $R_i^A(x, Q^2) < 1$ for $x \lesssim 0.1$: the shadowing region.

At LHC energies, most of the contribution to the hadron production cross-section originates from partons in the shadowing region, as represented in Fig. 2.7.

Fig. 2.8b shows the comparison of the gluon parametrisation for lead at $Q^2 = 10\,\mathrm{GeV}^2$ in the EPPS16 [75] and nCTEQ15 [76] analyses. For this case, both parametrisations are in agreement but they have very different central values and uncertainties for different x values.

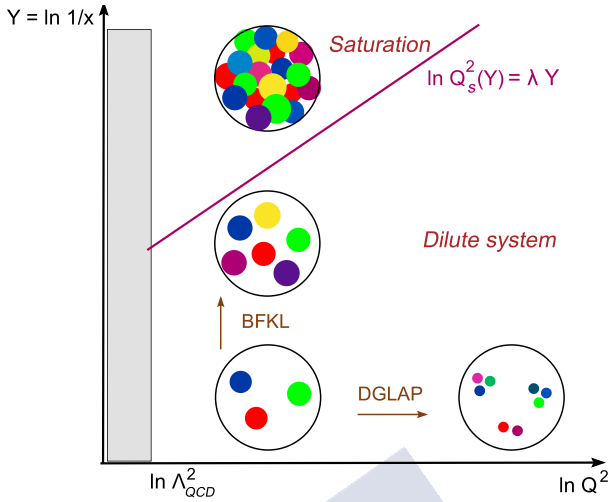


Figure 2.9: Evolution of the partonic density for DGLAP (fixed x, $Q^2 \to \infty$) and for BFKL (fixed x, $Q^2 \to \infty$), showing the eventual saturation of the gluon density at $Q_s(x)$. Figure taken from Ref. [84].

2.4.2 Saturation and Colour Glass Condensate

At low values of x the gluon density increases strongly, as shown in Fig. 2.2. The rapid rise of the gluon distribution at fixed Q^2, $x \to 0$ and $s \to \infty$ is given by the BFKL equation [78–80]. Experimentally, this behaviour in the proton PDF was seen first by HERA [81]. For the stability of the theory it is required that gluons reach a maximal occupation number at a given saturation momentum $Q_s(x)$ [82, 83]. This saturation phenomenon is illustrated in Fig. 2.9. For the case of nuclei, saturation is reached at higher values of x as a result of the Lorentz contraction of the nuclear parton density in the probe rest frame [73].

In this regime, the QCD dynamics are non-linear and non-perturbative, and can be described by the Colour Glass Condensate (CGC) effective field theory [85,86]. The CGC theory can be applied to pA collisions to predict particle production cross-sections for kinematics dominated by the low x contribution. In general, the models based on this approach predict a suppression of charged particle production in pA collisions relative to pp, supporting a strong shadowing at very low x. For inclusive charged particle production at $\sqrt{s_{\mathrm{NN}}} = 5\,\mathrm{TeV}$, several predictions in the CGC framework are available [87–89].

2.4.3 Cronin effect

In 1975, a enhancement of the inclusive hadron production at high p_{T} ($p_{\mathrm{T}} \gtrsim 1.5\,\mathrm{GeV}/c$ in proton-nucleus (pA) collision was observed with respect to scaled pp collisions [90]. Similar enhancements were found later at central pseudorapidity at RHIC [70,91,92], and also, but less pronounced, at LHC [41, 69]. The enhancement usually occurs for values of

$p_{\mathrm{T}} \gtrsim 1.5\,\mathrm{GeV}/c$, and tends to decrease with p_{T}. For the case of charged hadron production, protons seem to have stronger enhancement than pions and kaons [93, 94].

Explanations for this effect are usually made in terms of multiple interactions in the large x regime, and the object experimenting the re-scattering can be incoming or outgoing partons [95,96]. The contribution to the charged particle cross-section from multiple scattering can be calculated using pQCD techniques similar as those presented in Sec. 2.1.1, but including additional partonic processes. Recent calculations [97] can explain the enhancement for muons from heavy-flavour hadron decays at LHC and RHIC [98, 99] and for inclusive hadron production at RHIC [71]. Nevertheless, there is not an unambiguous explanation for the Cronin effect yet. Alternative approaches to explain the Cronin effect are based on final-state recombination of soft and shower partons [100], and they are able to reproduce RHIC data.

2.5 Objective and definition of the observables

The experimental approach to test CNM effects is to compare observables in proton-lead collisions with the corresponding quantities obtained in pp collisions. One privileged magnitude is the nuclear modification factor for charged particles ($R_{p\mathrm{Pb}}$). The nuclear modification factor is defined as

$$R_{p\mathrm{Pb}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}}) = \frac{1}{A}\frac{\mathrm{d}^2\sigma^{\mathrm{ch}}_{p\mathrm{Pb}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}}{\mathrm{d}^2\sigma^{\mathrm{ch}}_{pp}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}}, \tag{2.11}$$

where $A = 208$ is the number of nucleons in the lead ion and $\mathrm{d}^2\sigma^{\mathrm{ch}}_{p\mathrm{Pb},pp}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})/\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}$ are the differential cross-sections for charged particles in proton-lead collisions and proton-proton collisions, respectively. The subscript "cms" indicates the pseudorapidity measured in the nucleon-nucleon centre-of-mass system of the collision. The laboratory system must be inferred in the following when no subscript is noted. As a fundamental input for the nuclear modification factor, the measurement of these production cross-sections is the goal of the present analysis. They can be obtained as

$$\frac{\mathrm{d}^2\sigma^{\mathrm{ch}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}} = \frac{1}{\mathcal{L}}\frac{N^{\mathrm{ch}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{\Delta p_{\mathrm{T}}\Delta\eta_{\mathrm{cms}}}. \tag{2.12}$$

Here, N^{ch} is the number of prompt charged particles produced in a given range of $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$, Δp_{T} and $\Delta\eta_{\mathrm{cms}}$ are the size of those bins, and $\mathcal{L}$ is the integrated luminosity value of the corresponding data set. The crucial part of this analysis is to calculate the prompt charged particle multiplicity distributions with respect to η and p_{T}. The luminosity of each dataset is also an important input, more details about the luminosity determination can be found in Secs. 4.1.1 and 4.2.1.

In addition to the nuclear modification factor, the backward-to-forward ratio is defined

as

$$R_{\mathrm{BF}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}}) = \frac{\displaystyle\int_{-b}^{-a} \frac{\mathrm{d}^2\sigma^{\mathrm{ch}}_{p\mathrm{Pb}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}}\mathrm{d}\eta_{\mathrm{cms}}\mathrm{d}p_{\mathrm{T}}}{\displaystyle\int_{a}^{b} \frac{\mathrm{d}^2\sigma^{\mathrm{ch}}_{p\mathrm{Pb}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}}\mathrm{d}\eta_{\mathrm{cms}}\mathrm{d}p_{\mathrm{T}}}, \tag{2.13}$$

where $[a, b]$ is a given η_{cms} interval. The *backward* (*forward*) region corresponds to the region with negative (positive) η_{cms}, taking as positive the direction of the proton beam. The R_{BF} equivalent to the ratio between the $R^{\mathrm{ch}}_{p\mathrm{Pb}}$ in symmetric forward and backward regions, and measures the relative modification between both acceptances. This magnitude does not require the *pp* cross-section measurement.

Similarly, the forward-to-backward ratio can be defined as

$$R_{\mathrm{FB}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}}) = \frac{\displaystyle\int_{a}^{b} \frac{\mathrm{d}^2\sigma^{\mathrm{ch}}_{p\mathrm{Pb}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}}\mathrm{d}\eta_{\mathrm{cms}}\mathrm{d}p_{\mathrm{T}}}{\displaystyle\int_{-b}^{-a} \frac{\mathrm{d}^2\sigma^{\mathrm{ch}}_{p\mathrm{Pb}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}\eta_{\mathrm{cms}}}\mathrm{d}\eta_{\mathrm{cms}}\mathrm{d}p_{\mathrm{T}}}. \tag{2.14}$$

The goal of this work is to measure the transverse-momentum dependence of the prompt charged particle spectra in *p*Pb collisions and in *pp* collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$, to determine the production cross-sections using Eq. 2.12 and then $R_{p\mathrm{Pb}}$, R_{BF} and R_{FB} with Eqs. 2.11, 2.13 and 2.14. As it will be explained in Sec. 5.3.1, *p*Pb measurement covers a pseudorapidity range of $-5.3 < \eta_{\mathrm{cms}} < -2.5$ and $1.6 < \eta_{\mathrm{cms}} < 4.3$, where η_{cms} represents the pseudorapidity measured in the nucleon-nucleon centre-of-mass system. The *pp* measurement spans over $2 < \eta_{\mathrm{cms}} < 4.8$.

There is no previous determination of charged particle spectra in the forward and backward regions in proton-lead collisions at the LHC and, therefore, this study will provide invaluable information for the understanding of particle production in soft QCD and for the tuning of Monte-Carlo generators. In addition, these measurements allow to make the first determination of the nuclear modification factor for charged particles in the backward and forward regions. This will allow CNM effects to be tested in a broad range of x of $10^{-6} \lesssim x \lesssim 10^{-4}$ and $10^{-3} \lesssim x \lesssim 10^{-1}$, according to Fig. 2.7.

2.5.1 Definition of prompt charged particle

The differential cross-sections for charged particles do not consider all final-state charged particles produced in a given *pp* or *p*Pb collision but only prompt (or primary) charged particles. Therefore, to fully specify the main observables it is necessary to establish an unambiguous definition of "prompt charged particle". For this, the recommendations of the LHC Minimum Bias and Underlying Event Working Group are followed [101,102]: a "prompt charged particle" is *any hadron or lepton, with a mean lifetime* $\tau > 0.3 \times 10^{-10}$ *s, produced directly in the collision or from decays of shorter-lifetime*

particles. In the following, any mention to "charged particle" or "prompt charged particle" will refer to this definition. Effectively, this definition includes π^-, K^-, p, e^-, μ^- and their antiparticles and several baryonic resonances known as hyperons which containing the s quark (Ξ^-, Σ^+, Σ^-, Ω^- and their antiparticles) if they are produced promptly. For the rest of the document, the particle and its complex conjugate must be understood, unless specifically mentioned.

Apart from identifying the prompt charged particles that meet the previous definition, it is important to detail which neutral particles meet the lifetime requirement. Their daughter particles are not considered as prompt, so they are a background in the analysis (see "secondary particles" in Sec. 5.3.2.3 for more details). These are baryons (n, Λ, Ξ^0 and their antibaryons), mesons (K^0_S and K^0_L) and stable particles such as photons and neutrinos.

In Fig. 2.10, the relative abundance of the different prompt charged particle species is shown as produced by different MC generators. For pp (Figs. 2.10a and 2.10b), the generator is Pythia [103] with a specific LHCb configuration [57], while for Pbp (Figs. 2.10c and 2.10d) and for pPb (Figs. 2.10e and 2.10f) EPOS is used, considering the LHC model [58]. The kinematic acceptance is limited to the detector coverage. For all cases, the most important contribution are pions ($\sim 80\%$), followed by kaons ($\sim 13\%$) and protons ($\sim 7\%$). The presence of muons and electrons is merely testimonial. The relative particle abundances show very little dependence with η, both for PYTHIA and EPOS generators. A more detailed study of particle abundances is made in Fig. 5.4.3.

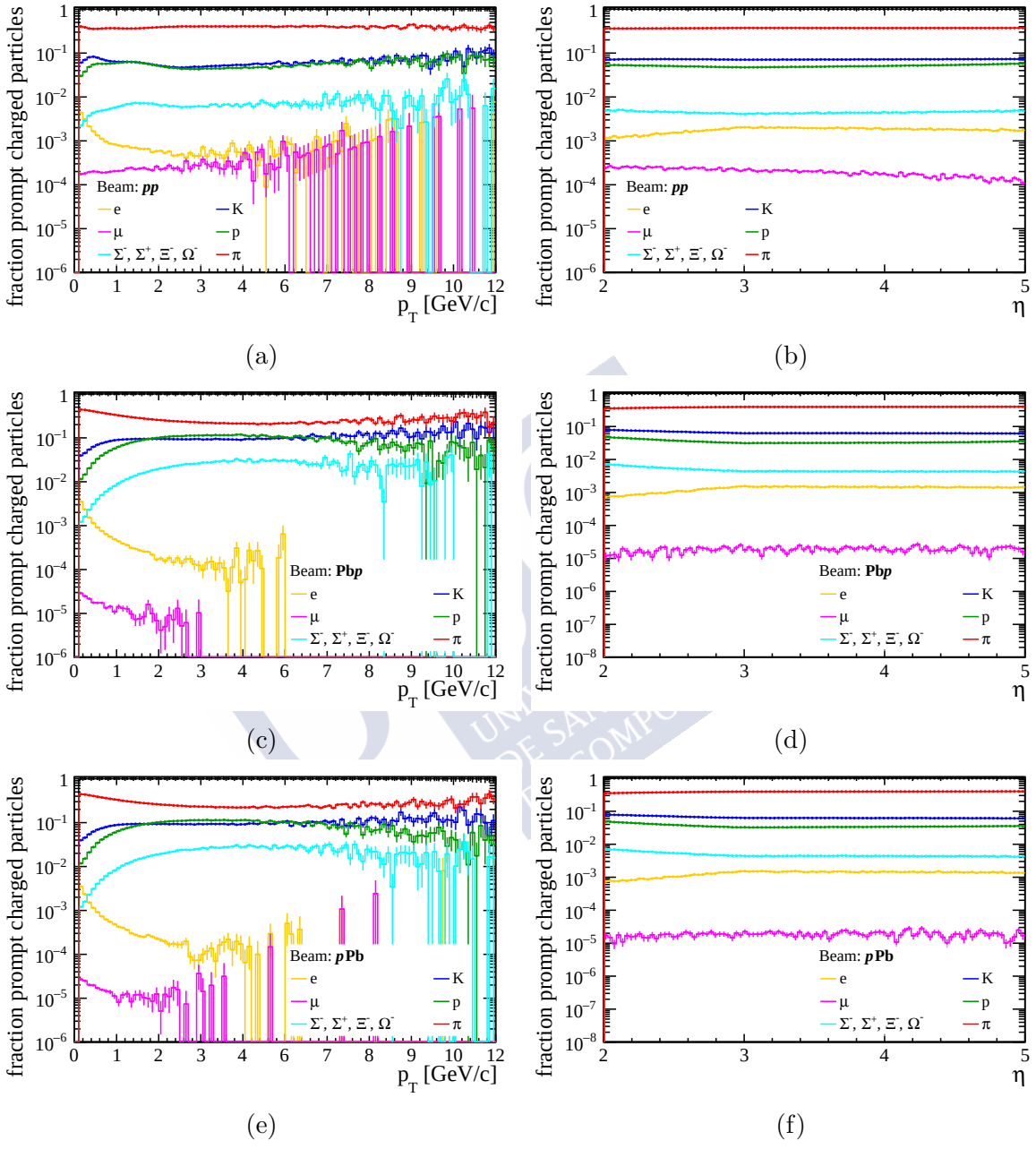


Figure 2.10: Relative abundance of the different particle species within the LHCb acceptance as a function of p_{T} (left) and η (right). Figures are prepared with simulated samples generated with EPOS for proton-lead and Pythia for pp. Figs. 2.10a and 2.10b (top) correspond to pp events, Figs. 2.10c and 2.10d (centre) to Pbp and Figs. 2.10e and 2.10f (bottom) to pPb events, respectively.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3

The LHCb experiment at the LHC

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest particle accelerator in the world, built at the CERN (from the French, *Conseil Européen pour la Recherche Nucléaire*) accelerator complex close to Geneva, Switzerland. The machine is installed in a 26.7 km tunnel constructed in the eighties initially to host the Large Electron-Positron (LEP) collider. The tunnel is located in-between the Swiss and French border and lies between 45 m and 170 m below the surface. A detailed description of the LHC machine can be found in Ref. [104].

The purpose of the LHC is to perform hadronic collisions at unprecedented high energies. In the standard configuration, the machine accelerates two counter-rotating beams of protons in two separate rings. The circular trajectory of the beams is kept with superconducting magnets that produce a magnetic field above 8 T. These magnets are placed inside a cryostat of superfluid He at 1.9 K to ensure their proper performance. The refrigeration of the magnets is a critical aspect of LHC operation, as well as the preservation of the ultra-high vacuum inside the tubes where the beams travel. The maximum laboratory energy per proton is 7 TeV, and when the beams collide the centre-of-mass energy ($\sqrt{s}$) of the colliding proton pairs is up to 14 TeV.

Before the injection in the LHC, a proton beam passes sequentially through the following preaccelerators: LINAC 2 (replaced in 2020 by the LINAC 4), Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). At each stage, the beam increases its energy up to 450 GeV at SPS. Then, the beam is split in

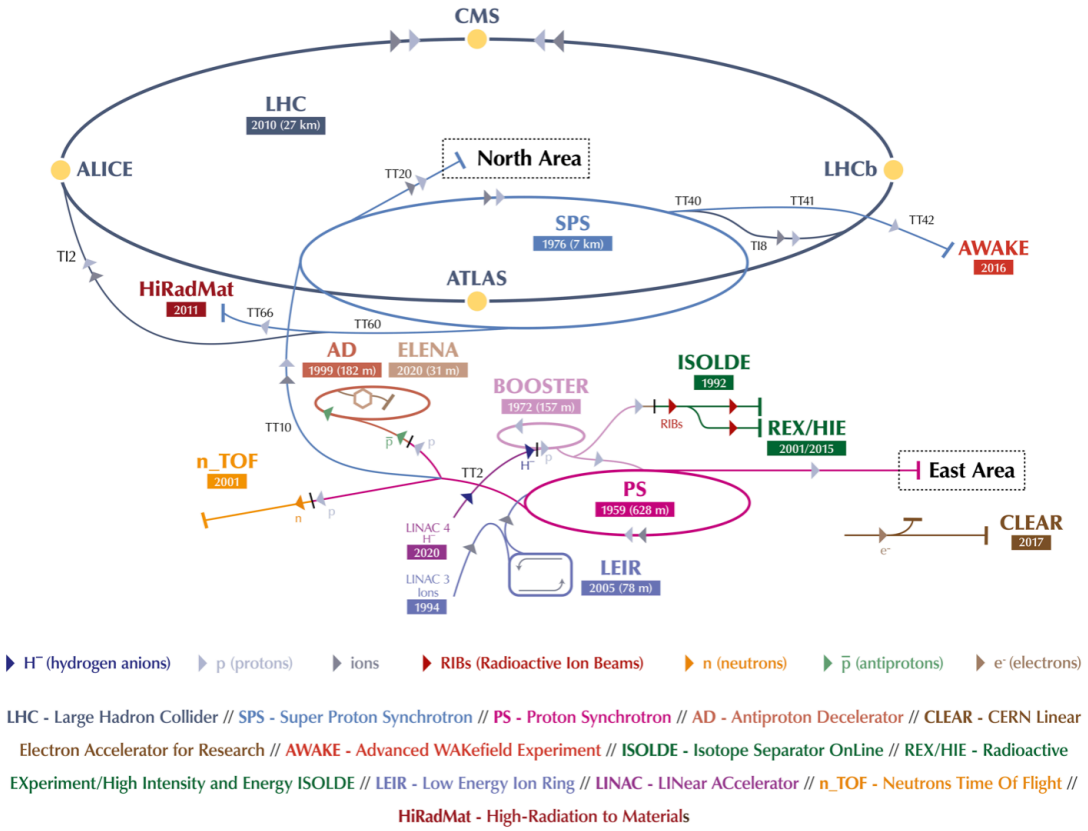


Figure 3.1: Diagram of the CERN accelerator complex in 2019. Note that the new LINAC4 substitutes the LINAC 2, which was operating during Run 1 and Run 2. Figure from [105].

two and enters the LHC, where both beams are accelerated up to their final energy before collision. Apart from protons, the LHC also operates with heavier ions, generally lead ^{208}Pb but also ^{129}Xe for a brief period in 2017. These ions are accelerated in a similar way as protons, although they start at the LINAC 3 facility where they are injected at the Low Energy Ion Ring (LEIR) and then to the PS to continue the same path as protons. An additional running mode, which was not initially foreseen in the LHC design, results when combining a proton beam with an ion beam to perform proton-ion collisions. The flow of protons and ions in the different facilities at CERN can be followed in Fig. 3.1.

The hadrons in the LHC beams are distributed in small packages, named bunches, and are organised along the LHC rings. The arrangement of the filled and empty bunches inside the LHC is referred to as the *filling scheme*. When operating in proton mode, the bunch spacing can be up to 25 ns resulting in a maximum collision frequency of 40 MHz. The maximum instantaneous luminosity at the LHC in this configuration is $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

On the contrary, the target peak luminosity in lead-lead runs is $\sim 10^{27}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, mostly because the LHC is populated with less bunches in this configuration.

The beams are slightly deviated at four points of the LHC to induce hadronic collisions. The tunnel has large caverns at these points where the main LHC experiments are located. These are: ATLAS (A Toroidal LHC Apparatus), CMS (Compact Muon Solenoid), ALICE (A Large Ion Collider Experiment) and LHCb (Large Hadron Collider beauty).

ATLAS [106] and CMS [107] are two general-purpose detectors (GPDs) that share similar scientific goals. Both investigate a wide range of physics, from studying the properties of the Higgs boson to searching for extra dimensions or particles that could make up dark matter. Their most remarkable achievement has been the discovery of the Higgs boson in July 2012 [108, 109]. Most of the detection capabilities of these detectors cover the central region, which corresponds to a small range around $\eta = 0$. This optimises the acceptance for the decay products of unstable heavy particles, which usually have large transverse momentum with respect to the beam.

ALICE [110] is designed to study the strong interaction in the QGP. As described in Sec. 2.2.2, it is necessary to collide heavy ions to achieve the required extreme energy densities. Hence, the ALICE detector is optimised for operation in AA collisions, corresponding to extremely high multiplicities. The ALICE acceptance is slightly particular as a result of the configuration of its detectors: a central barrel covering $\eta \approx 0$ is equipped with tracking, particle identification and calorimetry; and then, a forward muon spectrometer which spans for $2.5 < \eta < 4.0$. Additionally, there are several smaller detectors for global event characterisation and triggering in the forward region.

Finally, LHCb [4] is mainly oriented to the study of CP violation, flavour symmetry breaking and the search for new physics beyond the Standard Model through indirect searches. In this field multiple results can be highlighted, such as the recent evidence for the breaking of lepton universality in beauty-quark decays [111] or the first observation of CP violation in the charm sector [112]. Additionally, LHCb has evolved towards a much ambitious physics program including the study of heavy-ion collisions, as is demonstrated in the course of this thesis. Unlike the other detectors at the LHC, LHCb does not surround the interaction point but covers the low-angle or forward region. A more in-depth description of the LHCb experiment is given in the following (Sec. 3.2).

3.2 The LHCb experiment

The LHCb experiment [4, 113] is located at point 8 of the LHC, in the cavern where the previous DELPHI (Detector with Lepton, Photon and Hadron Identification) detector [114] operated during the LEP era.

The LHCb detector is a one-arm forward spectrometer with an angular coverage from 10 mrad to 300 mrad in the horizontal plane and from 10 mrad to 250 mrad in the vertical plane, corresponding to the pseudorapidity range of $2 < \eta < 5$. As mentioned

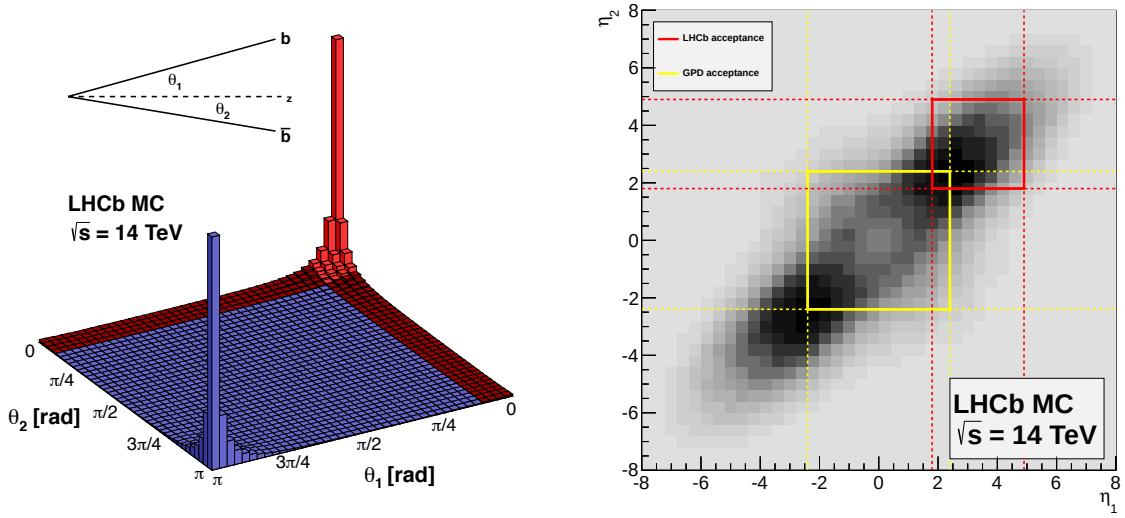


Figure 3.2: Left: Angular distribution of $b\bar{b}$ pair production in pp collisions at $\sqrt{s} = 14\,\text{TeV}$ [115]. Right: coverage in pseudorapidity of $b\bar{b}$ pairs for LHCb and a standard GPD, considering an acceptance of $1.8 < \eta < 4.9$ and $|\eta| < 2.4$ respectively. Figures from [115].

before, the forward coverage is the main distinctive feature with respect to the other LHC experiments. The reason to cover this kinematic region is to optimise the acceptance for B-hadrons which represent the core of the physics program at LHCb. Such hadrons originate in the hadronisation of a b or $\bar{b}$ which are mainly produced at low angles in proton-proton collisions at LHC energies. Fig. 3.2 shows the angular distribution of $b\bar{b}$ pair production at $\sqrt{s} = 14\,\text{TeV}$ as simulated by Pythia, showing a clear preference for low angles.

A key aspect of most LHCb analyses is the correct identification of the Primary Vertex (PV), where the proton-proton (or proton-lead) interaction occurs and B-hadrons are produced, and the Secondary Vertex (SV), where the B-hadron decays. The mean flight distance of such hadrons is $\sim 10\,\text{mm}$, requiring a high resolution in the determination of such vertices.

The detector is optimised for low pile-up collisions with low multiplicity, and the instantaneous luminosity during proton-proton data taking is levelled to $2-5\times10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$ as shown in Fig. 3.3, lower with respect to that in ATLAS and CMS. The pile-up reduction has as consequence a minor number of events with overlapping pp interactions, which eases the PV and SV separation. In addition, the average detector occupancy is lower in low pile-up conditions, allowing to relax some of the detector specifications such as the granularity and the radiation endurance. It is important to note that the mean particle multiplicity per solid angle is much higher in the forward region than in the central region, since most produced particles have a strong boost in the forward direction.

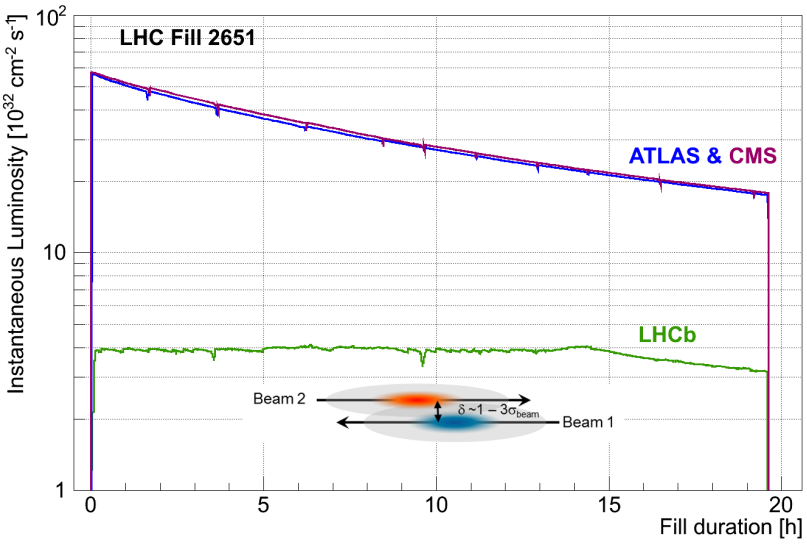


Figure 3.3: Instantaneous luminosity during LHC fill 2651 for ATLAS, CMS and LHCb experiments. The LHCb luminosity is kept in a range of 5% around the target $4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ during the fill by adjusting the transversal beam overlap. Figure from Ref. [113].

For most of its physics program, LHCb relies on the reconstruction of complex decay chains with many particles in the final state [116, 117]. A high reconstruction efficiency of the individual particle states is of paramount importance to ensure sizeable samples of decays to measure the observables of interest. In addition, the momentum resolution, specially of charged particles, has to be excellent to be able to clearly separate different resonances that decay to the same final states using their invariant mass distribution. This is the case of the B^0 and B_s^0 resonances [118] or the χ_{nc} excited states [119, 120]. The degradation of the momentum resolution usually originates in multiple scattering of particles with the detector material; therefore, the way to enhance the momentum resolution is to reduce the material budget to the minimum. This key principle permeates through all the LHCb design, and also ensures a high reconstruction efficiency.

In Fig. 3.4 a transverse view of the spectrometer is shown, with a representation of the main detection systems. By order of appearance from left to right, the different subdetectors are: the Vertex Locator, the RICH1 (Ring Imaging Cherenkov detector 1), TT (Tracker Turicensis), the magnet, three tracking stations (T1, T2, T3), RICH2, a muon station (M1), the calorimeter system formed by the Scintillating Pad Detector (SPD), the Preshower detector (PrS), the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL), and the rest of the muon system stations (M2, M3, M4 and M5).

Since 2013, the LHCb experiment also participates in the heavy-ion physics program at LHC. This decision was made to exploit certain characteristics of the experiment that

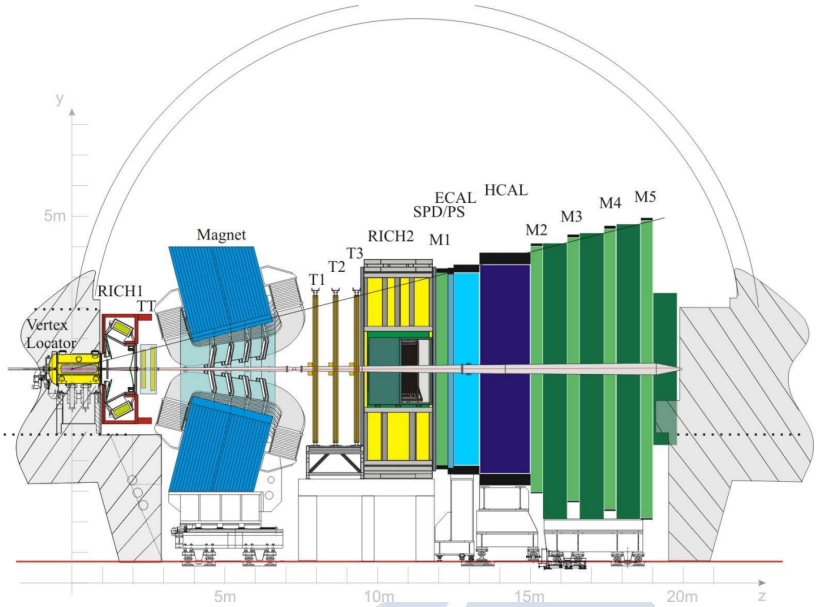


Figure 3.4: General layout of the LHCb detector. Figure from Ref. [4].

are of great use in this branch of physics. The most relevant of these features is the kinematic range which covers the forward region, providing an acceptance not reached by the other LHC detectors. As explained in Secs. 2.2.2 and 2.4, one of the core points of the heavy-ion physics research is the characterisation of CNM effects in pA collisions and the partonic structure of heavy nuclei. LHCb, with its forward acceptance, can probe very low x values with a variety of observables, as shown in Fig. 2.7. The ALICE detector has forward coverage as well but is limited to its muon spectrometer, restricting the physics reach in that region to resonances decaying to muons which are mostly quarkonium states. Instead, LHCb has full hadron reconstruction and identification capabilities in the region, as well as precise vertexing to separate prompt from detached particle production and a better momentum resolution. The main challenge for LHCb appears in very high multiplicity events. In lead-lead collisions, the operation is limited to $\approx 60\%$ most central events due to a saturation of the tracking system. Nevertheless, proton-lead collisions are fully within the LHCb multiplicity reach.

It is worth noting that the LHCb detector used for this thesis has been dismantled and is being replaced by an upgraded detector [121]. The main goal of such upgrade is to make possible the acquisition of data at higher instantaneous luminosity in proton-proton configuration of about $2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, using a very flexible fully software-based trigger. To achieve this goal a full replacement of the tracking system, RICH detectors, the muon system and the full readout electronics are required, among other improvements. The LHCb Upgrade is currently scheduled to start data-taking in the LHC Run 3 in 2022.

In the following sections, a more in-depth description of the different LHCb subsystems and their performance is provided. A special focus on the VELO and the tracking

system is made in Sec. 3.2.1 and Sec. 3.2.2, respectively, given that the present work relies mainly on the reconstruction of charged particles. The particle identification detectors (PID) will be briefly summarised in Sec. 3.2.3. Then, an overview of the the trigger system and strategy is presented in Sec. 3.2.4. To conclude this chapter, the LHCb simulation framework, which is a critical element in every LHCb analysis, is presented in Sec. 3.2.5.

3.2.1 Vertex Locator

The Vertex Locator [122] (VELO) is a silicon microstrip detector located surrounding the beam pipe region where the collisions occur. The VELO provides precise measurements of the coordinates of tracks originated from charged particles. This track information allows for the precise detection of the PV, the point where hadrons from both beams interact. Additionally, the decay point of unstable particles, like b- and c-hadrons produced in the collision, is usually a few cm away from the PV. The high resolution of the VELO is able to resolve these positions to identify particles produced in the primary interaction or in secondary decays. This also permits accurate measurements of the decay lifetimes and the impact parameter of particles.

The VELO is a crucial detector for the LHCb operation. Apart from being used to find PVs and SVs, tracks reconstructed at the VELO are used as seeds for the rest of the tracking system in track reconstruction, as will be explained in Sec. 3.2.2.4. The VELO is also a key ingredient of the high-level trigger, which will be described in Sec. 3.2.4. Information from the VELO, such as the number of reconstructed tracks or the number of found vertices, is generally used as luminosity counters, which are reference quantities proportional to the luminosity, and are fundamental for its precise determination [123].

The VELO consists of 21 detection stations, each formed by two semicircular microstrips silicon sensors, of type R and of type ϕ, which cover the radial and the azimuthal coordinate respectively. The detector stations are arranged along the beam line over a length of around one metre as shown in Fig. 3.5. The radius of each sensor is approximately 42 mm and the thickness is around 300 µm. At the centre of each sensor there is an 8 mm opening to allow the LHC beam circulate unimpeded. Type R sensors measure the radial coordinate via azimuthal strips. Type ϕ sensors measure the angular coordinate using radial strips. Charged particles produced in the collisions traverse the silicon detectors generating electron-hole pairs; the produced current is detected using application-specific electronics.

When operating, the inner part of the sensors is only 7 mm from the LHC beams, which is smaller than the aperture required by the LHC during injection. To address this issue, the VELO is mounted in two retractable halves that are closed when stable-beam conditions are declared. The VELO sensors are located in a secondary vacuum, separated from the primary vacuum of LHC where the primary collisions occur by a RF-box. The region facing the beam, known as RF foil, is only 0.3 mm thick, to minimise the degradation of the momentum measurement of particles due to multiple scattering. The material of the VELO is known to great detail [124], and is implemented in the LHCb

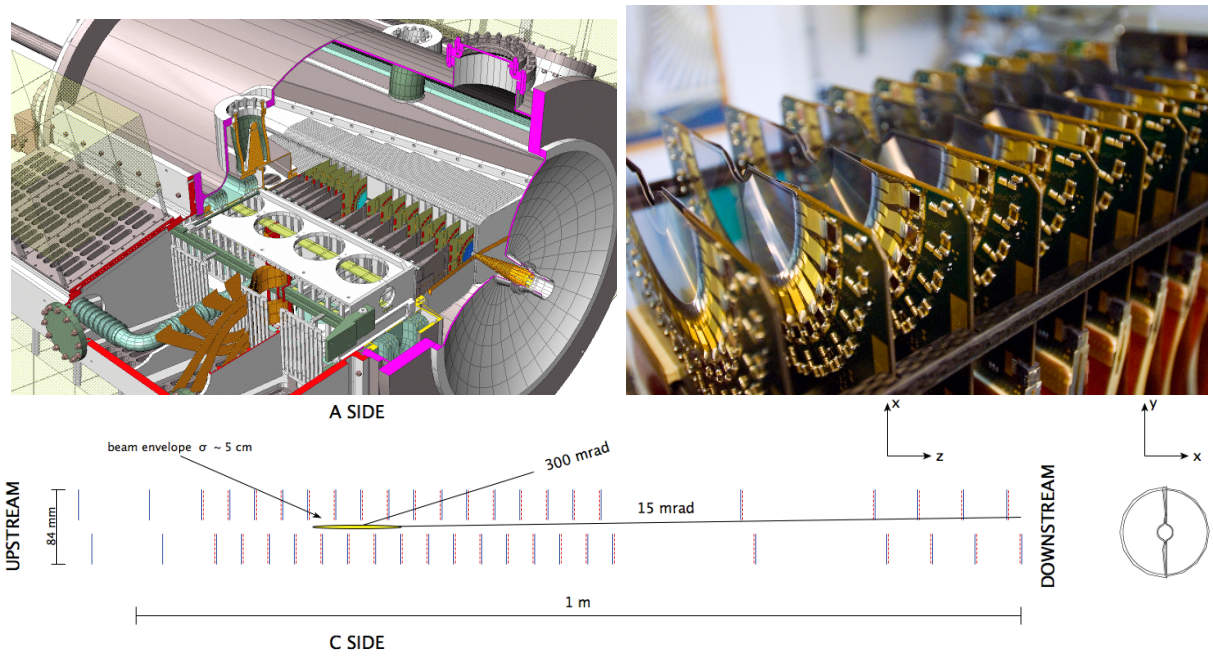
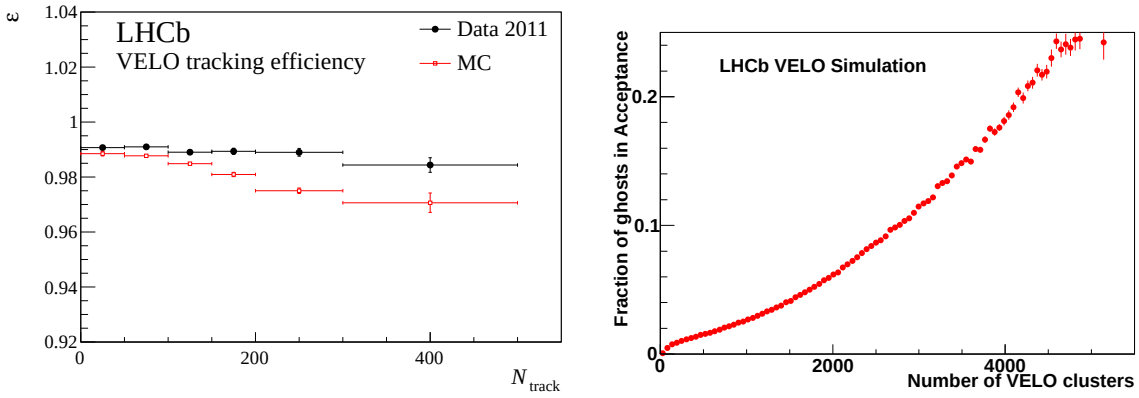


Figure 3.5: Top left: a cut-away view of the VELO sensors (blue) with their retractable support structure and the LHC vacuum. Top right: photography of one side of the VELO sensors. Bottom: disposition of the VELO sensors along z and cross-section of a VELO sensor in the xy plane. Figures extracted from Ref. [122].

simulation framework (see Sec. 3.2.5).

Regarding the performance of the VELO, the sensors have a signal to noise ratio of approximately 20 and a best hit resolution of 4 µm is achieved at an optimal track angle. A PV resolution of 13 µm in the transverse plane and 71 µm along the beam axis is achieved for vertices with 25 tracks, and an impact parameter resolution of less than 35 µm is achieved for particles with p_{T} greater than 1 GeV/c. The track reconstruction efficiency is typically 98% or higher in the data, as it is shown in figure Fig. 3.6a. Other relevant parameter is the fraction of ghost tracks, defined as a track in which less than 70% of the VELO clusters of the track are from a simulated particle. This corresponds to in 0.5% for randomly triggered events and 1% from triggered events. The fraction of ghosts in VELO increases with the detector occupancy as shown in Fig. 3.6b, where the total number of clusters at VELO, $N_{\mathrm{VELO}}^{\mathrm{clusters}}$, in the event is consider to compute the occupancy. These numbers are for the standard operation in pp collisions. For proton-lead operation, the detector occupancy reaches higher values, and for those events a degradation of the performance is expected.



(a) (b)

Figure 3.6: Fig. 3.6a: VELO tracking efficiency in data and simulation from 2011 as a function of the number of reconstructed tracks in the event, N_{track}, indicating the error bars the statistical uncertainty. Fig. 3.6b: fraction of ghost tracks at VELO as a function of the number of VELO clusters. Figures extracted from Ref. [122].

3.2.1.1 Primary vertex reconstruction

The reconstruction of PVs at LHCb is performed using VELO tracks [125]. Since the VELO is far from the magnet, there is almost no magnetic field and tracks are straight lines. The reconstruction algorithm interprets the track distribution in a event to infer the origin point of most tracks. Two contradictory requirements exist when optimising the PV reconstruction: a high efficiency, understood as the ratio between reconstructed and generated vertices in the simulation, and the rate of false PVs. False PVs can originate from random association of tracks in high multiplicity events. In general, high multiplicity PVs, where more tracks are produced, are easier to reconstruct as more tracks have the chance to be in the VELO acceptance. The difficulty increases in events with multiple PVs.

The reconstruction is performed in two steps: seeding and fitting. In the first step the PV candidates are selected by searching for the space points where an accumulation of track trajectories may be observed. In the second step the weighted least square method is employed to find the final vertex position. The candidates are fitted by descending multiplicity, which is useful to prevent the low multiplicity secondary vertices from absorbing tracks from their corresponding PVs. After fitting, the separation between the obtained PV and the already fitted ones is checked to protect against false PVs. The seeding and fitting steps are repeated until there are no new PVs in the event.

3.2.2 Tracking system

The tracking system of the LHCb detector aims at determining the trajectory of charged particles produced in the interaction and to measure their momentum. Sev-

eral subsystems contribute to the track reconstruction: the VELO (already discussed in Sec. 3.2.1), the magnet, the silicon tracker (ST) and the outer tracker (OT). They will be summarised in the following.

3.2.2.1 Magnet

A warm dipole magnet is in charge of creating a magnetic field in between the upstream and downstream tracking stations to curve the trajectory of charged particles allowing their momentum coordinates to be determined. The total magnetic field produced is of 4 T m. The polarity of the magnet is periodically changed during data taking to study possible systematic uncertainties in the measurements, specially in those sensitive to the charge of the signal, such as the determination of CP asymmetries.

The magnetic field integral, $\int B\mathrm{d}l$, created by the magnet is known with a relative precision of 10^{-4} and the position of the B-field peak with a precision of few millimetres, to achieve the target momentum resolution. An array of Hall probes is installed to perform scans of these parameters. The magnetic field is precisely measured in all LHCb regions.

3.2.2.2 Silicon tracker

The silicon tracker covers the region closer to the beam pipe. It consists of two detectors based on silicon microstrips: the Tracker Turicensis (TT), located upstream the magnet, and the Inner Tracker (IN), present in the inner region of the tracking stations T1, T2 and T3. Both TT and IT use silicon microstrip sensors with a strip pitch of about 200 µm. Fig. 3.7 shows a representation of the tracking stations.

The main function of the TT is linking VELO tracks with those produced in the tracking stations downstream the magnet. The TT is located just before the dipole magnet, and it is a 150 cm wide and 130 cm high planar tracking station which covers the full acceptance of the detector. The IT is located downstream the magnet, consisting in three subdetectors covering the region closer to the beam pipe of T1, T2 and T3. The IT layers are 120 cm wide and 40 cm high. Each four ST stations have four detection layers. The layers are arranged in a stereo layout, the external layers having vertical strips, whereas the internal layers have strips oriented with a $\pm 5°$ angle with respect to the vertical.

3.2.2.3 Outer tracker

The Outer Tracker (OT) [126, 127] covers the largest fraction of sensitive area in stations T1-T3. It is made up of straw-tube drift chambers, which are arranged in a array of straw-tube modules with two staggered layers of drift-tubes with inner diameters of 4.9 mm. In each station, modules are disposed in four layers, in a similar arrangement as the ST layers. They cover a large active area of $5971 \times 4850\,\mathrm{mm}^2$.

The diagram of the detector is shown in Fig. 3.7. Note how the region closer to the beam is covered by the IT. This is because the density of particles per angle is larger

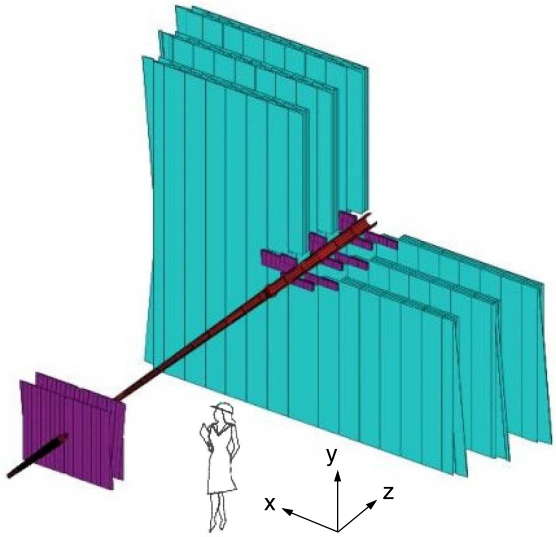


Figure 3.7: Diagram of the tracking stations as located in the beam pipe. The ST stations (TT and IT) are represented in violet, while the OT is represented in cyan. Figure taken from Ref. [4].

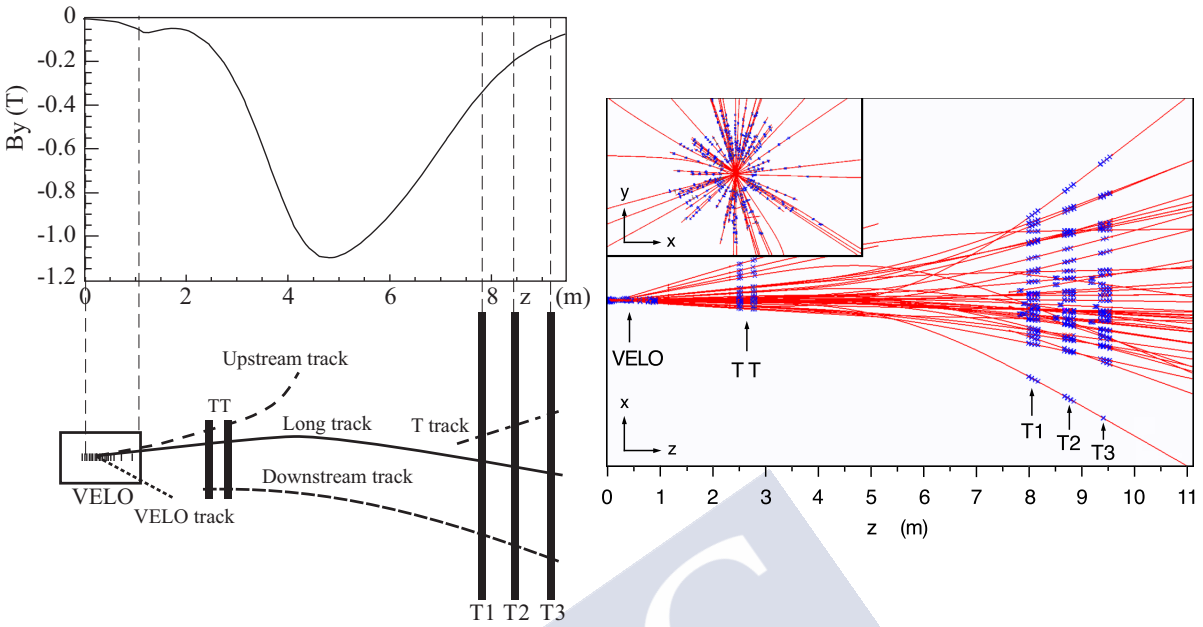
when approaching the beam line, and the resolution and radiation hardness requirements cannot be covered with the specifications of drift chambers.

3.2.2.4 Track reconstruction at LHCb

The LHCb reconstruction software takes into account information from the different LHCb tracking detectors and the map of the magnetic field to determine the charged particle tracks and their momentum. This is implemented with an algorithm that pairs hits in the different tracking stations to reconstruct the tracks. Track reconstruction at LHCb is embedded in the LHCb event reconstruction application called Brunel.

Tracks can be classified by the origin of the hits used to reconstruct them:

- **VELO track.** They are made up only from VELO hits and do not have hits from other stations. Typically, they have large angle or are backward tracks.
- **T track.** They only have hits in the T stations.
- **Upstream track.** They traverse VELO and TT stations. They are generally low momentum tracks removed from the acceptance by the magnetic field. They have poor momentum information but are useful for RHIC1 calibration, flavour tagging and for reconstructing specific decays.
- **Downstream track.** These only consider hits in the TT and the tracking stations after the magnet. In many cases, they are produced by the decay products of resonances with lifetimes enough to escape the VELO vessel, like K^0_S and Λ.



(a) (b)

Figure 3.8: Fig. 3.8a: scheme with the different kinds of tracks provided by the reconstruction software (bottom) and magnetic field profile (top). Fig. 3.8b: display of the reconstructed tracks in an event and the assigned hits in the xz plane. Figures from Ref. [4].

- **Long track.** They traverse the full tracking system from the VELO to the T stations, optionally also have hits on TT. They are optimal for the majority of physics studies because they have the most precise momentum determination.

The mentioned track types are represented in Fig. 3.8a. The reconstruction procedure starts with a search for track seeds in the VELO region [128] and the T stations where the magnetic field is low as shown in Fig. 3.8a. These initial track candidates are combined with hits in the T stations using different algorithms [129,130]. After long track candidates have been found, their trajectories are refitted with a Kalman filter [131], which considers corrections from multiple scattering and energy loss due to ionisation. The quality of the fit is given by the χ^2 per degree of freedom of the fit, χ^2/ndf. Fig. 3.8b shows the reconstructed tracks in a typical event.

After successful reconstruction, tracks are stored in the `TES["Rec/Track/Best"]` location of the DST. A DST is the data format that keeps event information, and is used as input for the LHCb analysis software known as DaVinci. There, only the "best" tracks are stored, *i.e.* if a VELO track is matched with a T track to form a long track, only the long track will appear.

The tracks that do not correspond to the trajectory of a real charged particle are called fake or ghost tracks. Most of these fake tracks originate from wrong associations

between VELO tracks and tracks in the T stations as a result of the large distance between both stations. From Fig. 3.4, this distance is more than 5 m. A precise determination of fake tracks is of paramount importance for the analysis of this thesis, so they will be studied in more detail in Secs. 5.3.2.1 and 6.4.1.

The efficiency is measured using a tag-and-probe technique with $J/\psi \to \mu^+\mu^-$ decays [132]. One of the muons is used as "tag" which is fully reconstructed, while the other is used as "probe", and is only partially reconstructed. After reconstructing J/ψ candidates with an invariant mass fit, a search for a long track match for the partially reconstructed probe is made. If found, the candidate is marked as efficient. The final efficiency is obtained with the ratio of efficient candidates over the total number of candidates. The technique can be used to measure precisely the tracking efficiency in data and simulated events to correct for possible discrepancies. The results from this calibration procedure are used in the present analysis, and is described in Sec. 6.2.

3.2.3 Particle identification system

Particle identification (PID) consists of classifying the detected particles according to their mass. The identifiable particles at LHCb are γ, e, μ, $\pi^\pm$, $K^\pm$ and p. The PID system comprises the RICH detectors, the calorimeters and the muon system. The analysis of this thesis is made on an inclusive sample of charged particles, thus no PID information is necessary to classify the signal candidates. However, the analysis exploits PID information in the studies of the background from secondary particles addressed in Sec. 6.4.2.

All the information from these detectors is usually integrated in a combined likelihood in order to maximise the identification efficiency and to minimise the misidentification rate. The efficiency varies between $90 - 100\%$ for different particles. In analyses, the efficiencies and the misidentification rates are usually computed using calibration samples [133, 134], since the LHCb simulation does not reproduce well those quantities. In the following, the three PID systems are briefly discussed.

3.2.3.1 RICH system

The LHCb RICH system [135] consists of two different detectors: RICH-1, located between the VELO and TT, and RICH-2, between T3 and M1. Their primary role is the identification of charged hadrons (π, K, p), but they also contribute in the identification of charged leptons (e, μ) specially at low p. Their working principle relies on the production Cherenkov light rings by charged particles passing through a radiator gas with a specific refractive index n. The radius of the produced rings allows the Cherenkov angle θ_C to be measured, providing the mass of the particle through $\cos\theta_C = 1/(n\beta)$, where β is the relative speed of the particle relative to the speed of light, and the momentum measurement in the tracking system. In LHCb three radiator gases are used: aerogel and fluorobutane (C_4F_{10}) at RICH1, and CF_4 at RICH2. The use of two Cherenkov detectors with different radiators provides full coverage in the momentum range $2 - 100\,\text{GeV}/c$.

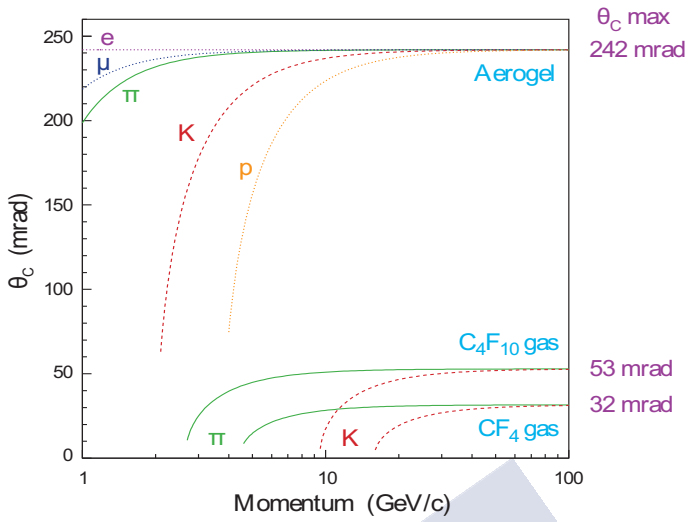


Figure 3.9: Reconstructed Cherenkov angles as a function of the track momentum in the different RHIC radiators. Figure taken from Ref. [4].

Fig. 3.9 shows the dependence of the Cherenkov angle with the particle momentum for each radiator.

The RICH 1 covers the low and intermediate momentum region, $2-40\,\mathrm{GeV}/c$ over the full spectrometer angular acceptance of $25-300\,\mathrm{mrad}$. The acceptance is limited at low angle by the size of the beam pipe upstream the magnet. The RICH 2 covers the high-momentum region $15-100\,\mathrm{GeV}/c$, over the angular range $15-120\,\mathrm{mrad}$. The Cherenkov photons emitted by charged particles traversing the RICH radiators are focused into ring images on the photon detector planes, situated outside of the spectrometer acceptance. A hybrid photo detector (HPD) specifically designed for RICH detectors is used. This HPD employs vacuum tubes with a 75 mm active diameter, with a quartz window and multialkali photocathode. The generated photoelectrons are focused onto a silicon pixel array of pixel size $2.5 \times 2.5\,\mathrm{mm}^2$, using an accelerating voltage of $-16\,\mathrm{kV}$. A total of 484 HPDs are close-packed to cover the four photodetector planes. The tubes feature a high detection efficiency. Both detectors have achieved a high resolution in the Cherenkov angle determination.

3.2.3.2 Calorimeter system

The Calorimeter System [136] is structured in several subdetectors: the Scintillating Pad Detector (SPD), the Pre-Shower Detector (PrS), the Electromagnetic Calorimeter (ECAL) and the Hadron Calorimeter (HCAL). Each one of these subsystems fulfils a specific function and as a whole they enable the identification of electrons and hadrons as well as the measurement of their energy and location. The calorimeter system has also a major role in the fast hardware trigger L0 [137] by measuring the deposited transverse

energy E_{T}. The scintillation light generated in the whole calorimeter system is transmitted to PhotoMultipliers (PMTs) which transform this light into an electric signal. Each system and its specific function are summarised in the following.

1. **SPD and PrS.** These detectors are two walls of scintillator pads separated by a lead curtain with a thickness corresponding to 2.5 radiation lengths. Each SPD cell provides a binary information depending on the comparison of the deposited energy with a threshold to distinguish charged particles from neutrals. The energy is then measured in a PrS cell to contribute to photon-electron identification.

2. **ECAL.** This detector is located next to the PrS and is designed to measure showers of electrons and photons. Layers of lead are used as absorber, which are alternated with layers of scintillating material. It has a thickness of 25 radiation lengths to ensure the full containment of the high energy electromagnetic showers and to get an optimal energy resolution. The total energy deposited in the ECAL in an event is also used to determine the number of participating nucleons, also known as centrality, in collisions of heavy-ions.

3. **HCAL.** It is located next to the ECAL, and it provides a measurement of the transverse energy of hadrons both for the L0 trigger and for contributing to the offline PID. It has a thickness is 5.6 interaction lengths. Iron is used as absorber and the active material are scintillating tiles.

3.2.3.3 Muon system

The muon system [138] is composed of five stations M1-M5 placed along the beam axis with protective geometry. Stations M2-M5 are placed downstream the calorimeters and are interleaved by iron absorbers of 80 cm thick to select penetrating muons. Station M1 is located downstream the RICH2 station and before the calorimeters to improve the p_{T} measurement in the trigger, avoiding the disturbance from the multiple scattering in the absorbers. The detector technology includes multi-wire proportional chambers (MWPC) for the M2-M5 stations and a gas electron multiplier (GEM) for the M1 station.

The detection of muons is a crucial part of LHCb operation, both for fast triggering at L0 and for identifying and measuring the momentum of muons which have a prominent role in the study of quarkonia, rare decays and CP violation. The trajectory of tracks reconstructed in the tracking system is extrapolated to the muon stations searching for compatible hits. If found, the track is classified as a muon.

3.2.4 Trigger system

The LHCb trigger system uses the information collected by the different subdetectors to reduce the event rate, from the 40 MHz rate of bunch crossings at LHC, to about $3-5$ kHz, the maximum rate to write to storage. This reduction is achieved while keeping

all events which are interesting for physics analyses. The trigger is composed of two levels: Level-0 (L0) and High Level Trigger (HLT).

The L0 trigger is implemented in hardware and reduces the rate from the 40 MHz to 1 MHz. It includes information from the calorimeters, specially the SPD and PrS, the muon system and the pile-up system of the VELO. Generally, an event is kept if certain threshold conditions on the calorimeter clusters or the muon tracks are fulfilled. The pile-up system is shown in Fig. 3.5 (bottom) as the two VELO sensors at the most right side of the figure, and they are used to reduce the impact of high multiplicity events that would be otherwise selected due to large combinatorics.

The HLT is a software trigger running in computing nodes forming the Event Filter Farm (EFF). This trigger performs a more exhaustive selection based in a full event reconstruction, reducing the event rate from 1 MHz down to 2 kHz. As it is software based, it is more flexible than L0 and allows easier adjustment for different physics analysis. The trigger consists on two steps, HLT1 and HLT2. At HLT1, a partial event reconstruction is made to confirm the L0 decision and to identify objects of interest in the event. These are defined prior to data-taking in *trigger lines*. In general these are particular track configurations with certain p_{T} or geometrical requirements. At HLT2, a full event reconstruction is made, with a very close quality to offline reconstruction, which allows for a finer selection of the particular event topology than at HLT1.

The trigger conditions for a specific data-taking run are specified with the *trigger configuration key* (TCK). A given TCK encodes the full set of trigger lines, with the specific threshold values of the trigger requirements and prescales, that were active during data-taking. Each run has one unique TCK. The LHCb trigger framework is implemented in a software package known as Moore.

3.2.5 Simulation and data flow at LHCb

As was explained in Sec. 2.3, the use of Monte-Carlo generated samples is fundamental to conduct physics studies in a hadron collider. At LHCb, there is a standardised framework to produce high-quality simulated samples that reproduce reliably the data to perform fundamental studies such as estimation of detection efficiencies, determination of background contributions or assessment of the detector resolution, among others.

The LHCb simulation framework is called Gauss [139], and comprises a series of applications involved in the different steps of the simulation chain. Proton-proton collisions are usually generated with the Pythia event generator [52], while the event generator used for proton-ion collisions is generally EPOS-LHC [58]. The particularities of these two models were already presented in Sec. 2.3. The decay and time evolution of the produced particles is implemented with EvtGen [140]. The Photos package [141] is used to introduce final-state radiation in the simulated samples. After the generation of all final-state particles, the toolkit Geant4 [48, 142] is used to transport them through the experimental setup. The description of the detector geometry and data-taking conditions is adapted to those in place during the targeted data sample using the detector description

database (DDDB) and the condition database (CondDB).

The output of Gauss is passed to another framework called Boole [139]. Boole is in charge of digitasing the simulated data into the same format provided by the experiment electronics and the DAQ system. Then, the output can be processed with Moore (see Sec. 3.2.4) as during online data taking, and then with Brunel (see Sec. 3.2.2.4) to implement the event reconstruction. The output format, both for simulated and real data, is usually a DST. An extra step, called stripping, is performed to apply an extra layer of filtering to the data to optimise the data storage resources. The DaVinci software application is used for the stripping, and also to process data from the DST into the final format used for analysis. This format is generally a `.root` file of the Root analysis framework [143]. The description of the datasets for the present analysis is addressed in the next chapter.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4

Description of the datasets

In this chapter, information about the data and simulation samples used in the analysis is collected. The study relies on data acquired in 2013 of proton-lead collisions and a sample of proton-proton collisions acquired in 2015. The description of the proton-lead and proton-proton samples is addressed in Sec. 4.1 and Sec. 4.2 respectively. Simulation samples are available to perform a variety of studies, such as the reconstruction efficiency, the effects of the imposed selection requirements and the contribution from background tracks. The simulation samples are described in Sec. 4.3.

4.1 Proton-lead data

In this analysis, data from proton-lead collisions acquired during February 2013 are analysed, corresponding to a total integrated luminosity of $\mathcal{L} \approx 1.6\,\text{nb}^{-1}$. This dataset corresponds to LHC fills in the range [3510, 3544]. The average instantaneous luminosity during the data taking was $\mathcal{L} \approx 3 \cdot 10^{27}\,\text{cm}^{-2}\,\text{s}^{-1}$, lower that the nominal for pp collisions, and the maximum recorded peak interaction does not exceed $\mu \approx 0.04$, being μ the average number of visible interactions per bunch crossing. This feature can be checked in Fig. 4.1, where the average number of reconstructed primary vertices is shown for each run using a random trigger.

In order to access both the proton and the lead fragmentation regions, data were taken in two different configurations. In the *forward configuration*, or pPb, the proton beam points towards the LHCb arm, whereas in the *backward configuration*, or Pbp, the lead beam does. Following the usual LHCb convention, the forward configuration is taken

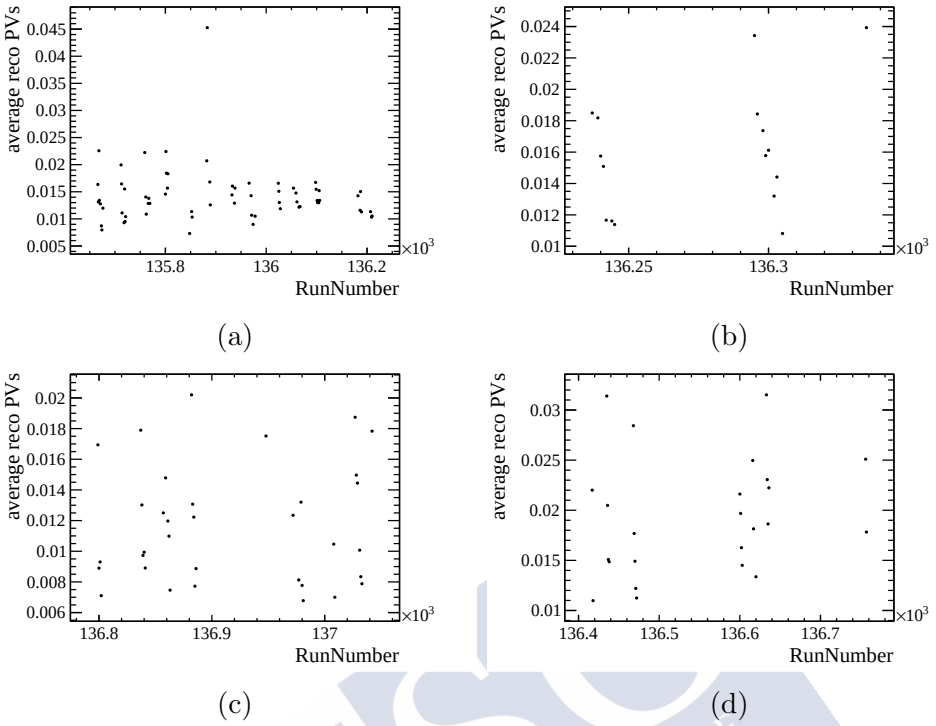


Figure 4.1: [
Average PVs in proton-lead, figure prepared by the author]Average number of reconstructed primary vertices per run for events with a random trigger (`Hlt1MBNoBias`) for the proton-lead data taking. Top figures correspond to pPb events (Fig. 4.1a to the MD configuration and Fig. 4.1b to the MU configuration) and bottom figures to Pbp events (Fig. 4.1c to the MD configuration and Fig. 4.1d to the MU configuration).

as positive values of pseudorapidity in the nucleon-nucleon centre-of-mass system, η_{cms}. For both configurations, the polarity of the magnet was switched, so that data sets with the polarity pointing upwards (*MagUp* or MU) and downwards (*MagDown* or MD) are available for both configurations. The integrated luminosity for each sample has been measured and can be checked in Sec. 4.1.1.

The energy per nucleon of the lead beam and the proton beam was 1.58 TeV and 4.00 TeV, respectively, which means an energy in the nucleon-nucleon centre-of-mass system of $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$. The asymmetry in the beam energies produces a boost of $y_{boost} \approx 0.465$ in the direction of the proton beam. Due to this boost, the LHCb kinematic acceptance in the nucleon-nucleon centre-of-mass system is different in the forward and backward configurations. More details about the kinematic coverage of the analysis are given in Sec. 5.3.1.

The trigger considered in the analysis is minimum bias. The minimum-bias trigger aims at not having any preferred event feature, selecting every bunch-crossing where a

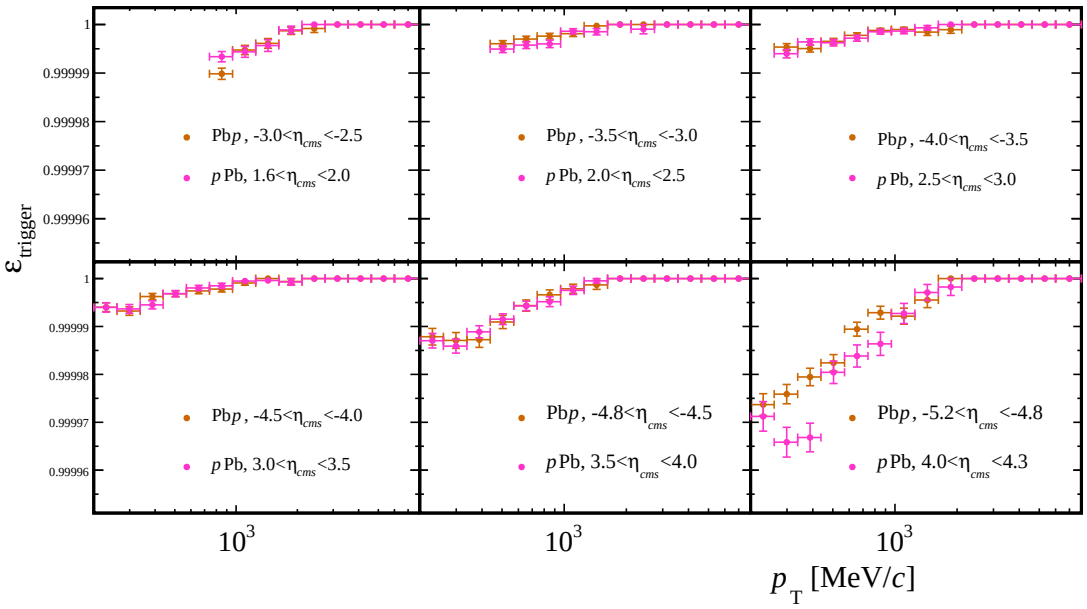


Figure 4.2: Trigger efficiency of the `Hlt1MBMicroBiasVelo` line from Pbp and pPb simulation

proton-lead (or proton-proton) inelastic interaction occurs. This can be achieved with minimal requirements to the activity in the detector. For pPb and Pbp events, the trigger line `Hlt1MBMicroBiasVelo` is considered. This line requires a single reconstructed VELO track to accept the event. The efficiency of this trigger line, defined as the number of prompt charged particles from events with `Hlt1MBMicroBiasVelo==True` divided by the total number of generated prompt charged particles, is close to 100% for all $(\eta_{\text{cms}}, p_{\text{T}})$, as shown in Fig. 4.2. A negligible drop in efficiency is seen for high η_{cms} and low p_{T}, which could be due to very low multiplicity diffractive events that do not leave tracks in the VELO acceptance. The hardware implemented L0-trigger was operated in minimum-bias mode forwarding all events to the high-level software trigger. Additionally, no particular HLT2 selection is required.

After data taking and storage, a stripping campaign was undertaken in order to optimise data usage. The stripping version where the minimum-bias data can be found is v20r3. Details on the considered stripping lines can be consulted in Tab. 4.1. For minimum-bias events the only difference between the sample before and after stripping is the introduction of prescaling factors. Their values are displayed in Tab. 4.1, along with the particular run number interval selected for each sample. These factors must be considered in the normalisation when computing the cross-section with Eq. 2.12, since the value for the integrated luminosity refers to the entire data sample considered in the analysis.

Several subsamples are discarded for the analysis due to various data quality criteria.

		Stripping Line	ODIN RunNumber	Prescale
Pbp	MD	`StrippingProtonIonMagDownMinBiasLine`	(136799 – 137045)	0.025
	MU	`StrippingProtonIonMagUpMinBiasLine`	(136417 – 136758)	0.110
pPb	MD	`StrippingIonProtonMagDownMinBiasLine`	(135576 – 136212)	0.080
	MU	`StrippingIonProtonMagUpMinBiasLine`	(136237 – 136341)	0.090

Table 4.1: Minimum bias lines from stripping v20r3 which are considered in the analysis. The information was extracted from [144].

	pPb		Pbp		pp
	MagDown	*MagUp*	*MagDown*	*MagUp*	*MagDown*
included TCK	0x006e004a 0x006d0048 0x006e0048	0x006e004a	0x006e004a	0x006e004a	0x0115014e
excluded TCK	0x006a1710 0x006e0049	0x006f004a	0x006f004a	–	–
excluded runs	135710, 135711, 135900, 135901			136729 to 136736	168072, 168128, 168183, 168182, 168143

Table 4.2: Included and excluded TCK and runs for every dataset.

Firstly, data acquired with TCK `0x006a1710` and also runs (135710, 135711, 135900, 135901) are excluded due to a muon inefficiency problem. Runs with TCK `0x006e0049` are removed due to higher `SumEt` (transverse energy E_{T} in the calorimeters) threshold at L0 with respect to the rest of the data sample. Additionally, TCK `0x006f004a` is not considered since the HLT1 line `Hlt1MBMicroBiasVelo` has a significant prescale factor of 1/100. The size of this sample is small compared with the rest of the dataset, which does not have any prescale in this line. These considerations follow the actions previously taken by LHCb analyses using this data-set (check [145], [146] or [147], for instance). Finally, runs (136729 – 136736) corresponding to the Pbp MU configuration are removed due to a displacement of the average y coordinate of the primary vertex (PV) distribution with respect to the rest of the dataset, as shown in Fig. 4.3. To summarise, Tab. 4.2 shows the accepted and rejected TCK and runs for each configuration.

4.1.1 Luminosity determination in proton-lead data

In proton-lead data, the luminosity is obtained from DaVinci, which extrapolates to the considered dataset the calibration of the LHCb Luminosity Working Group. A complete description of the calibration procedure for the present dataset can be found

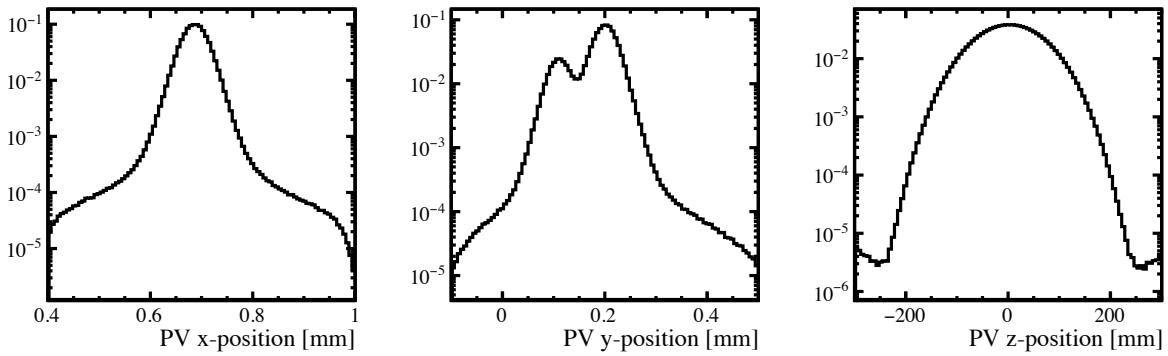


Figure 4.3: Spacial distributions of the reconstructed primary vertex for the Pbp MU data sample including the full dataset. Histograms are normalised to unity. A double-peak structure is observed for the y coordinate, due to a displacement of the beam identified in runs $(136729 - 136736)$. These runs were removed from the analysis.

Beam	Magnet	$\mathcal{L}$ [nb^{-1}]	relative uncertainty [%]	Stripping prescale
pPb	MD	0.849 ± 0.019	2.29	0.080
pPb	MU	0.1955 ± 0.0045	2.29	0.090
Pbp	MD	0.2564 ± 0.0064	2.5	0.025
Pbp	MU	0.2023 ± 0.0051	2.5	0.110

Table 4.3: Luminosity values for the proton-lead datasets.

in Ref. [123]. The recommended CondDB and DQFLAGS tags (`cond-20141002`) and (`dq-20140822`) are used, where the first refers to the detector condition database and the second to the data quality monitoring. The luminosity values used in the cross-section measurement are shown in Tab. 4.3. The luminosity must be multiplied with the prescale of the minimum bias stripping line detailed in Tab. 4.1.

4.2 Proton-proton data

Data from proton-proton collisions (pp) at $\sqrt{s} = 5.02\,\text{TeV}$ are used as reference for the measurement. These data were collected by LHCb in November 2015 with a 25 ns bunch spacing. Identified good-quality data correspond to four different fills: 4638, 4639, 4640, 4643. The corresponding run numbers and the number of analysed events are showed in Tab. 4.4. Due to considerations during the luminosity calculation, the following runs are removed from the analysis: 168072 (only run of fill 4638), 168128, 168183, 168182, 168143. No inversion of the magnet polarity was made during data taking, and therefore the entire dataset corresponds to the *MagDown* (MD) polarity.

Data are extracted from the `NoBias` stream, in particular from no bias events of collisions of leading bunches. There was only one trigger configuration during the com-

Fill	Runs	Number of leading bunch crossings
4639	168140, 168141, 168143, 168144, 168146	4
4640	168171, 168184, 168185, 168186, 168189 - 168196	9
4643	168233, 168234, 168237 - 168242, 168244 - 168249, 168261	23

Table 4.4: *pp* collision runs collected by LHCb at 5.02 TeV in 2015 within the `NoBias` stream. Only runs with dqflag='OK' and which are valid for luminosity are included.

plete data taking, corresponding to the `0x0115014e` TCK. Within this configuration, the HLT1 trigger line that selects no bias events from leading bunch interactions is `Hlt1NoBiasLeadingCrossing`, which is the one used in this analysis. Every event of the dataset also passes the `Hlt2NoBiasLeadingCrossing` selection. As opposite to the trigger line used in proton-lead, no requirements on the topology of the event are made. Events from leading bunch crossings are free from spillover, so they are specially suited for this cross-section measurement. The spillover corresponds to contamination between two neighbour colliding bunches, for instance due to the existence of a charge remnant in one of the detectors, and might be important for a 25 ns bunch spacing such as the dataset under discussion.

The instantaneous luminosity provided by the LHC during this data taking is higher in comparison with that of proton-lead collisions, and consequently, μ is higher in *pp*. Figs. 4.1 and 4.4 display the average number of reconstructed PV per run for a random trigger in proton-lead and proton-proton data takings, respectively. The figures show that the rate of reconstructed primary vertices is below 0.045 in proton-lead, but goes as high as 0.6 for proton-proton samples.

The reconstruction version for this dataset is Reco15a. Here, a baseline selection for the tracks is applied. Specifically, a cut in ghost probability (`GhostP` as defined in Sec. 5.3.2) of `GhostP` < 0.4 and track quality of $\chi^2/\text{ndf} < 4$. Only tracks that meet these criteria are available in the DST. Note how this differs from proton-lead reconstruction, where no requirement in `GhostP` is applied at reconstruction level.

4.2.1 Luminosity determination in proton-proton data

For *pp* data, the luminosity of the full Run 2 *pp* dataset has been measured by the Luminosity Working Group using the log-zero method, as it is explained in detail in [148]. However, the `NoBias` 2015 5.02 TeV dataset contains only leading bunch crossings, and therefore the calculation must be restricted to these particular bunches. The luminosity can be obtained as

$$\mathcal{L} = \frac{\sum_{\text{BX}} \langle \mu_{\text{BX}} \rangle \cdot N_{\text{BX,NoBias}}}{\sigma_{\text{VDM}}}, \tag{4.1}$$

where $\langle \mu_{\text{BX}} \rangle$ is the average number of interactions per bunch crossing and per run, $N_{\text{BX,NoBias}}$ is the number of events triggered by the no-bias leading bunch crossing trigger

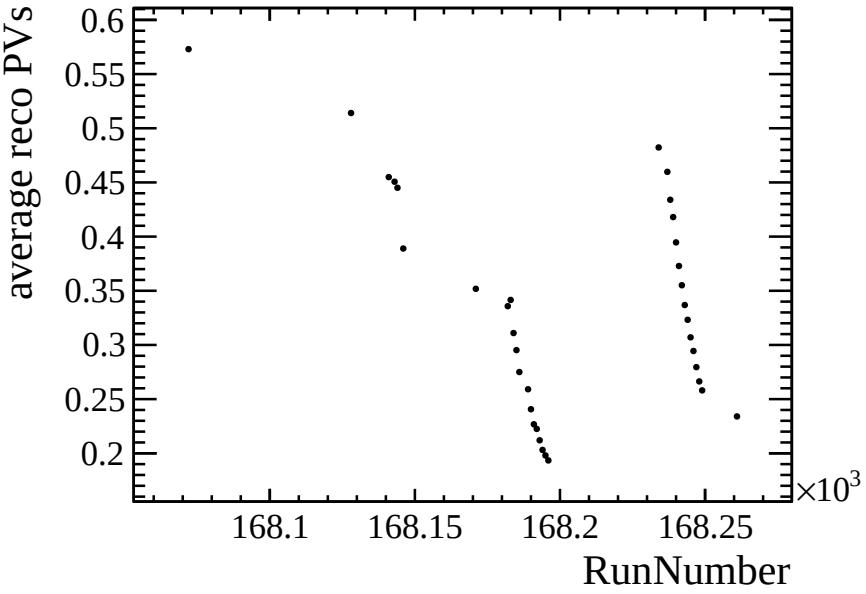


Figure 4.4: Average number of reconstructed primary vertices in the *pp* dataset per run for events selected by the `NoBias` trigger line `Hlt1NoBiasLeadingCrossing`.

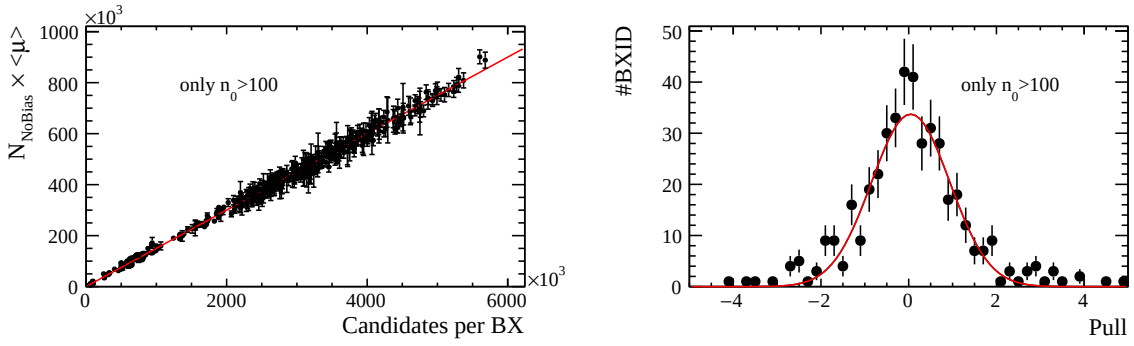
and σ_{VDM} is the reference cross-section for the used lumi counter[1] that is determined in a van der Meer scan.

The values of $\langle\mu_{\mathrm{BX}}\rangle$ have been obtained using rolling windows and are available per bunch crossing and run [149]. The value $N_{\mathrm{BX,NoBias}}$ is the number of events triggered by the `Hlt1NoBiasLeadingCrossing` and `Hlt2NoBiasLeadingCrossing` lines per bunch crossing and run. For the 2015 5.02 TeV `NoBias` dataset, every event stored in the FULL.DST corresponds to leading bunch crossings, so all of them fulfil both trigger lines. The value of σ_{VDM} for the vertex luminometer at 5.02 TeV is $\sigma_{\mathrm{VDM}} = 50.90696 \pm 1.0\,\mathrm{mb}$.

A few checks are performed to validate the luminosity result. The first cross-check can be done comparing with an alternative measurement of the luminosity for the 13 TeV early measurements `NoBias` dataset [150]. This dataset consists of collisions from leading bunch crossings as well. The values of $\langle\mu_{\mathrm{BX}}\rangle$ for that dataset are available as for the 5.02 TeV sample [149], and the number of $N_{\mathrm{BX,NoBias}}$ can be obtained by processing this dataset with DaVinci and considering events that pass the triggers `Hlt1MBNoBiasLeadingCrossing` and `Hlt2MBNoBiasLeadingCrossing`, which select collisions from leading bunch crossings.

Two particular runs were cross-checked: run 159978 from *MagUp* configuration and run 157584 from *MagDown*. The number of minimum-bias events $N_{\mathrm{BX,NoBias}}$ per bunch crossings is compared with the number reported in [150] (table 3 in the reference). The found values for the mentioned runs match exactly the values in the cited reference.

[1] A lumi counter is an observable proportional to the number of primary interactions which is used as reference to determine μ. The lumi counter used here is the number of reconstructed primary vertices.



(a) (b)

Figure 4.5: Fig. 4.5a: $\langle \mu_{\mathrm{BX}} \rangle \cdot N_{\mathrm{BX,NoBias}}$ with respect to the number of candidates for each bunch crossing in every run. The data are fitted to a linear function, shown in red. Fig. 4.5b: Pull distribution of $\langle \mu_{\mathrm{BX}} \rangle$ with the linear fit for each bunch crossing and run. The pull distribution is fitted to a gaussian function, shown in red.

With $N_{\mathrm{BX,NoBias}}$ and $\langle \mu_{\mathrm{BX}} \rangle$, the luminosity can be measured using Eq. 4.1, taking the measurement of σ_{VDM} for the vertex luminometer at 13 TeV, which is $\sigma_{\mathrm{VDM}} = 58.01925\,\mathrm{mb}$. The resulting luminosity for both runs agrees with the results from [150] at the 0.2% level. A more extensive test has been done finding agreement between the two results for all runs [149].

A second cross-check is to ensure that the measured luminosity for every bunch ID and run scales with the signal of the present analysis. The number of candidates to prompt charged particles, as defined later in Sec. 5.3, will be used as signal for this cross-check. This assumes that the corrections to the spectra, such as the reconstruction efficiency or the background subtraction, are uniform during the whole data taking, which is a valid assumption since no important variation in detector occupancy or (η, p_{T}) distributions for candidates has been observed. For more information, these corrections are discussed in detail in Chap. 6.

To verify that the scaling holds, the product of the numerator of equation Eq. 4.1, $\langle \mu_{\mathrm{BX}} \rangle \cdot N_{\mathrm{BX,NoBias}}$, can be plotted with respect to the number of candidates in each bunch crossing. This is shown in Fig. 4.5a, where the uncertainty on $\langle \mu_{\mathrm{BX}} \rangle$ has been estimated as

$$\delta\langle \mu \rangle \approx \sqrt{\frac{1}{n_0} - \frac{1}{n_{ev}}}, \tag{4.2}$$

where n_{ev} is the number of events from the lumi stream across all rolling windows in a particular bunch crossing and run and n_0 is the number of expected empty events from the lumi stream for a particular bunch crossing and run and can be obtained as $n_0 = n_{ev} e^{-\langle \mu \rangle}$. [2] Eq. 4.2 assumes the Poisson distribution, which is an approximation that is not expected to hold for bunch crossings from short runs. For this reason, only bunch

[2] Note that the subindex BX is omitted here for simplicity.

Fill	#Leading BX	Run	$N_{\texttt{NoBias}}$	$\mathcal{L}\,[\mu\text{b}^{-1}]$
4639	4	168141	365189	3.28835
		168144	844411	8.02087
		168146	519886	4.46605
4640	9	168171	8430544	63.0165
		168184	10320464	66.8026
		168185	14339610	86.6238
		168186	14427370	81.0045
		168189	14439336	74.7249
		168190	14437662	71.8792
		168191	13342156	62.7382
		168192	542007	2.82483
		168193	14431567	65.6735
		168194	14441732	61.0842
		168195	397908	1.69432
		168196	6399976	27.0506
4643	23	168234	14418506	145.141
		168237	11074417	101.476
		168238	5351617	47.6461
		168239	36957466	315.794
		168240	36991871	296.392
		168241	36998124	282.88
		168242	36990581	264.52
		168243	30468178	207.533
		168244	36952495	238.464
		168245	36997212	230.925
		168246	27303720	166.215
		168247	36938276	215.621
		168248	36992122	206.375
		168249	9047165	49.1201
		168261	8324742	40.7065
Total:			529486310	3489.70010568

Table 4.5: $N_{\texttt{NoBias}}$ and luminosity per run in the *pp* dataset. Empty cells have the same content as the cell above. The last row shows the sum of $N_{\texttt{NoBias}}$ and $\mathcal{L}$.

crossings with $n_0 > 100$ are included in this cross-check and in Fig. 4.5. This criterion excludes 31 out of 465 bunch crossings considered in the analysis. Data are fitted to a linear function which is shown in the plot. To assess if the spread is only due to statistical fluctuations, the pull distribution of the data with respect to the linear fit is shown in Fig. 4.5b. The distribution is symmetric and can be described with a gaussian, which is shown in red. The value of σ of the fitted gaussian is 0.892 ± 0.043. The observed variability can be then attributed to the statistical uncertainty of the procedure, which is negligible when considering the full sample. No important systematic biases are observed.

In Tab. 4.5, the value of $N_{\text{BX},\texttt{NoBias}}$ and the measured luminosity per run are presented. The total luminosity of the *pp* dataset is $\mathcal{L} = 3.490 \pm 0.070\,\text{nb}^{-1}$, considering a 2%

uncertainty. The value is about a 9% lower than the value computed with DaVinci and that was used previously in the analysis. This discrepancy is expected since the luminosity changes for the different bunch crossings and DaVinci does not scale properly when only leading bunch crossings are selected.

4.3 Simulation samples

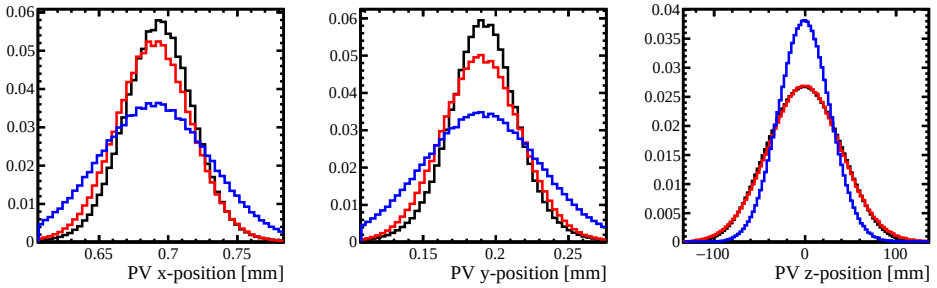
Along with the data collected with the LHCb detector, Monte-Carlo (MC) simulation samples for pPb, Pbp and pp events are used in this analysis. They have a crucial role to understand detector-induced biases and other experimental effects. A summary of the simulation samples employed in the analysis is presented in Tab. 4.6. The MC samples for pPb and Pbp have been produced with the official LHCb tune, using EPOS LHC [58] for the generation. Events are created with the minimum-bias configuration, and the number of proton-lead interactions in each event is fixed to one. Several productions with this configuration have been made during the analysis timeline. The first one (version sim09c) was produced by the end of 2017. In that version, there were inaccuracies in the settings of the beam parameters that led to a disagreement with data in the primary vertex position distributions. After tuning the parameters (check ref. [151] for the development), a second version (sim09d) was produced by mid 2018. The impact of this bug can be checked in Fig. 4.6, where the PV distribution is compared in data and simulation with versions sim09c and sim09e. In this figure, there is a clear mismatch in the width of some distributions between the first simulation version and data, in particular in the transverse coordinates for the Pbp sample and in all cases for the longitudinal coordinate z. The origin of the disagreement was found to be a misconfiguration of the beam option file. Finally, a bug affecting the generation of long-lived particles like K^0_S and Λ^0 was found (see ref. [152]) and a third version (MCsim09e) was produced.

The considered number of simulated pPb events is 5208647 (5226592) in the MD (MU) configuration, whereas the number of Pbp simulated events is 5165389 (5166396) in the MD (MU) configuration.

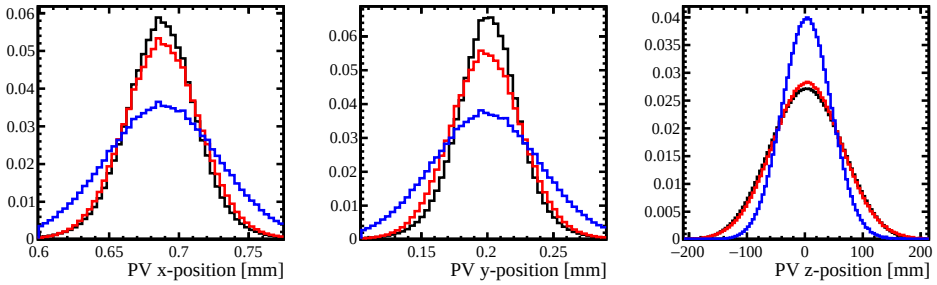
Prior to the availability of this official simulation, a private simulation sample generated with Hijing [153,154] had been produced. This sample consists of about 10^6 events generated both in the pPb and Pbp configurations, all with MD polarity.

The simulation sample of pp collisions consists of 10005820 events generated with Pythia in minimum-bias conditions with the LHCb official tune MCsim09d. As data are only available in MD configuration, only MD events were produced. The simulation is configured with an average number of interactions per bunch crossing $\nu = 1.5$, and therefore events with more than one pp interaction are included.

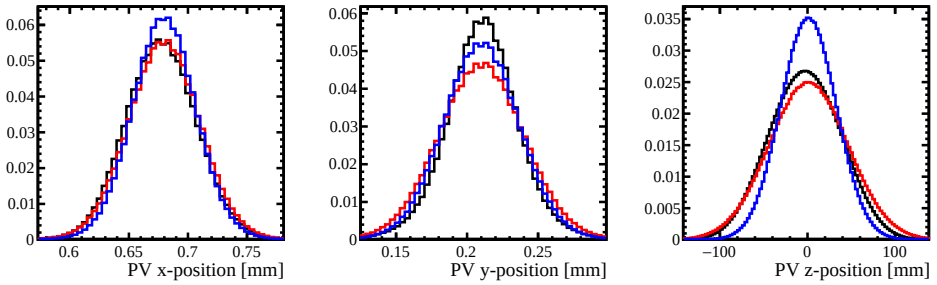
In addition, several samples of pp collisions have been produced to perform different cross-checks in the analysis. A sample of pp collisions was produced with the EPOS generator, simulating only the particle generation. This sample is used to study the relative particle abundances in pp data in Fig. 5.4.3 and Sec. 6.2.2. Also for this study, a



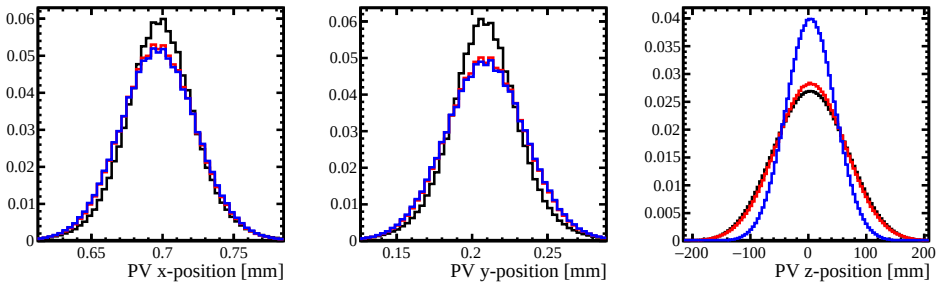
(a) Pbp MD



(b) Pbp MU



(c) pPb MD



(d) pPb MU

Figure 4.6: Comparison of the primary vertex position distributions for data (black) and simulation samples MCsim09c (blue) and MCsim09e (red). The vertical axis represents normalised entries of the primary vertex position.

Generator	Sim version	Beam	Magnet	#evts	Pile-up	Comments
EPOS	Sim09e	pPb, Pbp	MD, MU	$5 \cdot 10^{6\dagger}$	1 int/evt	*
Hijing	private	pPb, Pbp	MD	$1 \cdot 10^{6}$	1 int/evt	Private version used before EPOS available
Pythia	Sim09d	pp	MD	$1 \cdot 10^{7}$	$\nu = 1.5$	*
EPOS	private	pp	MD	$\sim 10^{7}$	1 int/evt	Only generator level.
Pythia	private	pp	MD	$\sim 10^{6}$	1 int/evt	Only generator level, with rope hadronisation.
Pythia	private	pp	MD	$\sim 10^{6}$	$\nu = 1.5$	Full simulation, no spillover.
EPOS	private	pp	MD	$\sim 10^{6}$	$\nu = 1.5$	Full simulation, no spillover.

Table 4.6: Simulation datasets available for the analysis.
† Requested number of events per configuration. The final numbers of produced events always slightly exceed these figures.
* Simulation datasets used to compute efficiencies and corrections.

sample of Pythia events using rope hadronisation tune [155] was produced.

Two additional samples of pp collisions have been produced, in this case following all the simulation and reconstruction steps. The first sample is made with Pythia and has the same options as the official MCsim09d but removing spillover effects. This was to verify the validity of the official simulation, since spillover effects should not be present in the data sample as it only includes interactions from leading bunch crossings as was explained in Sec. 4.2. The second is similar but the generation is done with EPOS. These samples are used in Sec. 7.3.2 to cross-check the procedure to estimate the fraction of secondary particles in data.

5

Preparation and selection of the data and simulation

£ In this chapter, the selection criteria for events and charged particle candidates are explained. The trajectories of charged particles, tracks, are measured by the LHCb tracking system as explained in Sec. 3.2.2.4. However, there are different background contributions that need to be studied. The idea of the selection is to minimise the influence of background events and background tracks, reducing the relative magnitude of corrections to the spectra. Additionally, an overview of control plots of the different datasets is presented. Agreement between data and simulation samples is studied as well.

5.1 Event Selection

An event corresponds to a bunch crossing determined by the LHC internal clock. There can be different bunch crossing types, depending on whether each bunch is filled with beam or not. Only events produced in bunch-bunch interactions are selected for the analysis, thus with `BunchCrossingType==3`. Other bunch crossing types, such as empty-empty (type 0), beam-empty (type 1) or empty-beam (type 2), can be used to study the features of background events.[1] The events are required to meet the minimum-bias trigger conditions that were detailed in Chap. 4. Additionally, not all of those events are considered for the analysis, but they are required to fulfil a series of criteria to

[1] Only proton-lead data contain information about these events. For proton-proton, only events from bunch-bunch crossings are available in the datasets.

suppress background events that contaminate the sample. These criteria are assessed in the following section Sec. 5.1.

Background. Apart from proton-proton or proton-lead beam collisions, other interactions may occur in the LHC beam pipe. Some of these processes can be misinterpreted by the trigger resulting in an accepted *background event*. In addition, proton-proton or proton-lead collisions can also be simultaneous to one of these spurious interactions. If they occur within the VELO acceptance, a legitimate primary vertex can be reconstructed in the same manner as for proton-proton and proton-lead collisions but the observed charged particle multiplicity would be contaminated with tracks originated in the parasitic interaction. This motivates the introduction of a requirement to suppress such events.

The main sources of background events are:

- *Beam-gas* interactions. These are produced when a colliding particle of one beam hits a gas molecule within the VELO vacuum.
- *Beam-splash* interactions. They occur when a particle from one beam interacts with material of the detector or accelerator support structure.

These two kinds of interactions share two key aspects: the available energy for particle production is less than in nominal bunch-bunch collisions, and the average interaction point is displaced from the nominal point for proton-proton and proton-lead collisions. Another background type that could arise is:

- *Interactions between a nominal and a previous bunch* (spill-over). However, these are mitigated in proton-lead data taking as the time between bunch crossings is larger than in the *pp* configuration. In *pp* data, this background is not present since only leading bunch crossings are considered.

Luminous region. The first strategy to reduce background events in proton-lead collisions is to exclusively consider events with a reconstructed primary vertex (PV) within a fiducial *luminous region*. In this way, background collisions are preferentially removed, as they spread over the entire VELO vessel. The size of the luminous region is defined with reference to the distribution of the PV coordinates (x_{PV}, y_{PV}, z_{PV}), being z parallel to the beam pipe and x and y the transverse coordinates. For each of them, the distribution is well described by a Gaussian core around the beam crossing location and a broader background contribution, as shown in Fig. 5.1. The mean value for each of the projections, $\mu_{x,y,x}$, and the standard deviation, $\sigma_{x,y,z}$, are determined by a Gaussian fit to the vertex distribution for each projection. The luminous region is defined within a range of three standard deviations around the respective mean values:

$$|x_{\mathrm{PV}} - \mu_x| < 3\sigma_x, \;\; |y_{\mathrm{PV}} - \mu_y| < 3\sigma_y, \;\; |z_{\mathrm{PV}} - \mu_z| < 3\sigma_z. \tag{5.1}$$

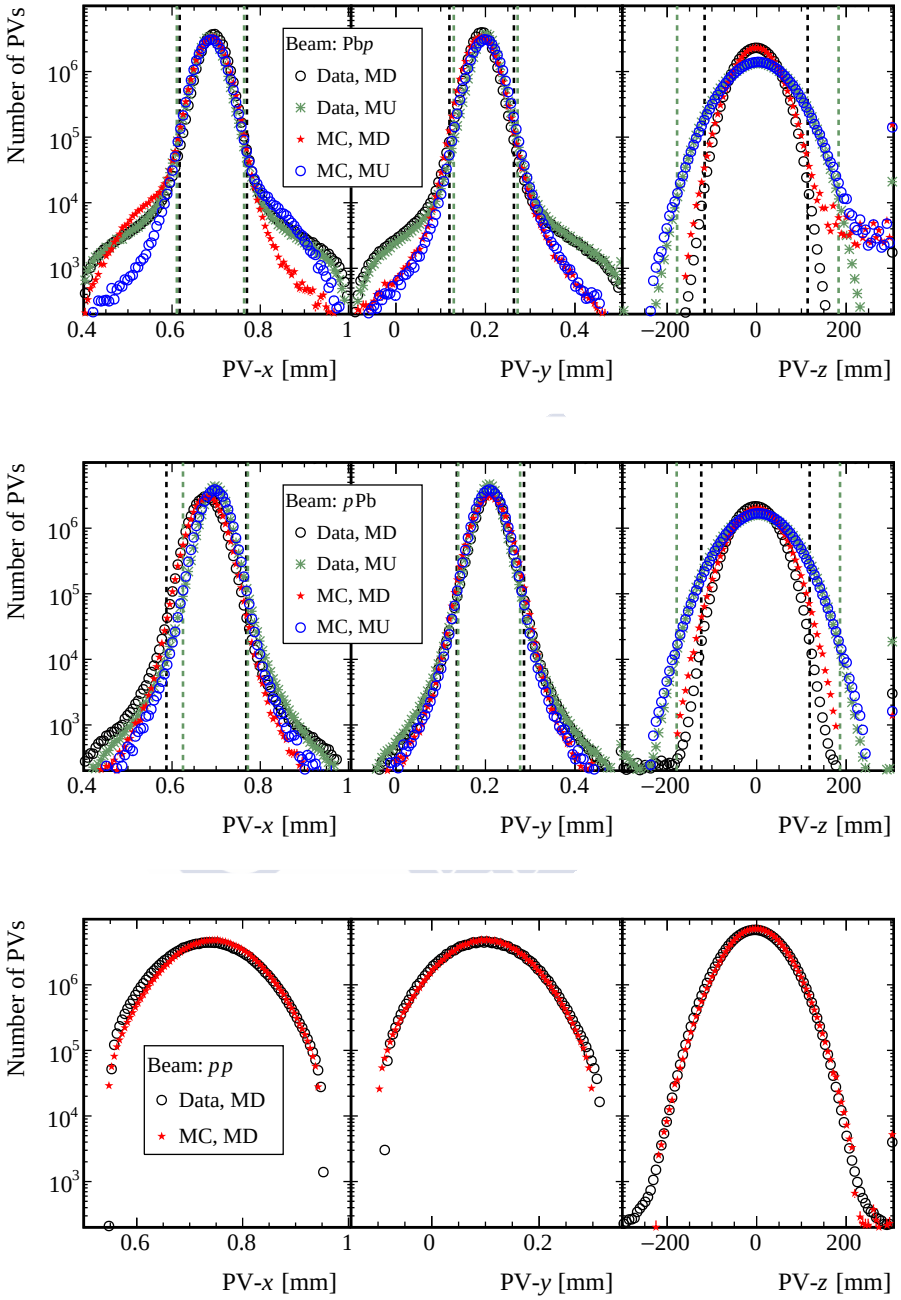


Figure 5.1: Spatial reconstructed primary vertex distribution for Pbp (top), pPb (middle) and pp (bottom) configurations. Black and green dots represent data, while red and blue dots represent simulation. Vertical axis are expressed in logarithmic scale, to ease the visualisation of the background. Simulation is scaled to the total number of PVs in data. Dashed vertical bars indicate the boundaries of the luminous region.

	pPb					
	MD			MU		
	x_{PV}	y_{PV}	z_{PV}	x_{PV}	y_{PV}	z_{PV}
μ[mm]	0.677	0.211	-2.7	0.698	0.208	3.83
σ[mm]	0.030	0.025	40.6	0.024	0.023	61.08

	Pbp					
	MD			MU		
	x_{PV}	y_{PV}	z_{PV}	x_{PV}	y_{PV}	z_{PV}
μ[mm]	0.694	0.192	-1.17	0.687	0.201	2.66
σ[mm]	0.025	0.024	38.5	0.025	0.024	60.40

Table 5.1: Result of the Gaussian fit used to define the luminous region for the forward (top) and backward (bottom) configuration for both magnet polarities.

The values for the parameters obtained from data are detailed in Tab. 5.1.

In Fig. 5.1, it can be seen how the simulation sample fails to reproduce the PV distribution outside the luminous region. This disagreement is not observed for the *pp* case (Fig. 5.1 bottom) where simulation and data distributions have very similar shape in the transverse coordinates. The comparison is not straightforward: proton-proton data and simulation samples include events with multiple primary interactions (pile-up), while in proton-lead data pile-up is very low. In proton-lead simulation only one primary interaction is generated per event. It is also interesting to notice that for the transverse coordinate x of the Pbp configuration (Fig. 5.1 top) there is an influence of the magnet polarity, since for MD the discrepancy is located in the right side of the plot while it is in the left side for MU. Nevertheless, none of these discrepancies have impact in the analysis as those events are removed by the luminous region requirement.

In Tab. 5.2, the fraction of kept events by the luminous region requirement is presented. When requiring a reconstructed PV within the luminous region, implicitly another requirement is imposed, which is that the event has at least one reconstructed PV. Therefore, the table presents the fraction of kept events after considering sequentially the following requirements: 1) the event has at least one PV, 2) at least one PV is within the luminous region, and 3), the event has exactly one PV and is within the luminous region. The application of requirement 3) is discussed in the following.

For the *pp* sample Fig. 5.1 shows that the impact of background events is less important with respect to proton-lead. For this reason, no restriction to the position of the primary vertex is considered.

Primary vertex condition. As stated in Chap. 4, data taking conditions for proton-lead and proton-proton samples were significantly different. One of the key differences is the average number of observed primary vertices, much higher in *pp* than in pPb and Pbp events (check Figs. 4.1 and 4.4). The origin of this difference is the higher value

Dataset	Data			Simulation		
	has reco PV	within lum. reg.	one PV	has reco PV	within lum. reg.	one PV
pPb MD	0.908	0.98	0.992	0.963	0.972	0.999
pPb MU	0.910	0.97	0.991	0.962	0.974	0.999
Pbp MD	0.937	0.98	0.983	0.960	0.974	0.989
Pbp MU	0.943	0.976	0.978	0.959	0.975	0.989

Table 5.2: Fraction of kept events in data and simulation samples after requiring: 1) that events have at least one reconstructed primary vertex, 2) at least one PV is within the luminous region and 3) that there is exactly one reconstructed PV and is within the luminous region. The fractions are incremental in the sense that they are computed with respect to the number of events that pass the previous criteria.

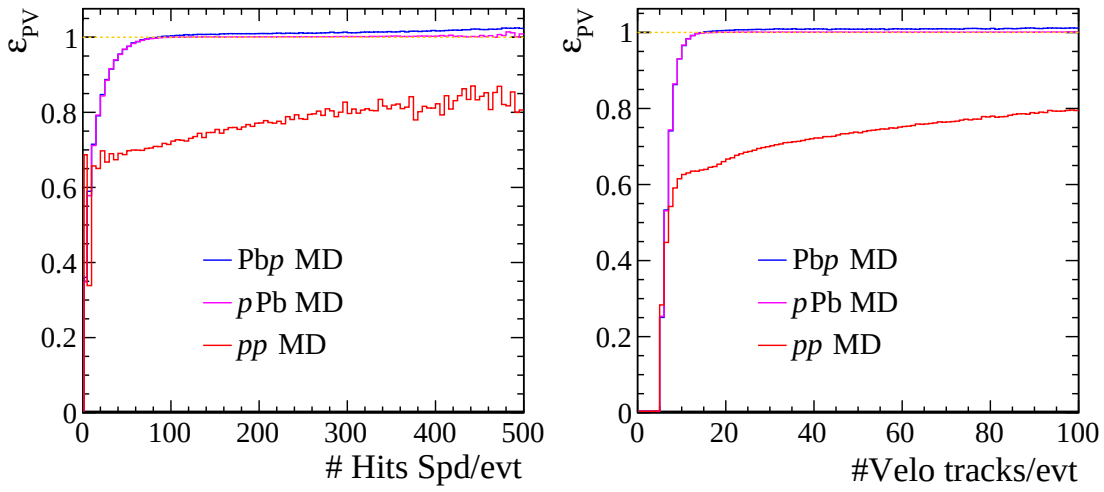
of the instantaneous luminosity in pp events as compared to proton-ion collisions. The low instantaneous luminosity in proton-ion collisions is very convenient since the detector would not be able to cope with a large pile-up of proton-lead collisions, which have much higher multiplicity than pp collisions. Thus, in proton-lead collisions, only a small fraction of the beam-beam bunch crossings will result in an inelastic interaction. However, due to the high charged particle multiplicity per primary vertex, these interactions will likely meet the trigger requirement of at least one reconstructed VELO track.

As shown in Fig. 5.2, the primary vertex reconstruction efficiency is higher in proton-lead events. In that figure the primary vertex reconstruction efficiency is represented as a function of the event multiplicity, which is accounted for with two variables: the number of hits in the SPD and the number of reconstructed tracks in the VELO. Note again that the dependency is similar with respect to the two variables due to their strong correlation.

The reason explaining the difference between proton-lead and proton-proton efficiencies is that the reconstruction of a PV requires at least three converging VELO tracks, which is fulfilled more easily in a high multiplicity environment. This can be checked in Fig. 5.2b, where the PV reconstruction cut-off at three VELO tracks is also evident. In addition, Fig. 5.2 shows that the reconstruction efficiency can be larger than 1 for high multiplicities in pPb and Pbp events. This is because, for some events, there are two reconstructed primary vertices instead of one. Note that in the proton-lead simulation exactly one interaction is generated for every event. Therefore, additional primary vertexes do not correspond to actual proton-lead interactions. The slight increase with multiplicity may imply that they correspond to an artefact of the PV reconstruction.

In Fig. 5.3, the total energy collected in the HCAL is represented with respect to the number of VELO clusters at different stages of the event selection. In addition to the selections already discussed, the possibility to restrict the number of reconstructed primary vertices to one in proton-lead is studied.

In these plots, the main diagonal band corresponds to nominal proton-lead collisions. In Figs. 5.3a and 5.3b, there is another band along the VELO clusters axis, which is reduced with the bunch-bunch requirement (Figs. 5.3c and 5.3d) and almost disappears with



(a) (b)

Figure 5.2: Primary vertex reconstruction efficiency for each beam configuration, pp (red), pPb (magenta) and Pbp (blue). The left plot shows the dependence with the multiplicity in SPD detector, while the right plot shows the dependence with VELO track multiplicity.

		pp	pPb	Pbp
Event selection	Trigger	`Hlt1NoBiasLeadingCrossing`	`Hlt1MBMicroBiasVelo`	`Hlt1MBMicroBiasVelo`
	Bunch crossing type	bunch-bunch	bunch-bunch	bunch-bunch
	number of PV	-	1	1
	PV position	-	within luminous region	within luminous region

Table 5.3: Event selection considered for the analysis.

the luminous region requirement (Figs. 5.3e and 5.3f). This band probably corresponds to beam-gas interactions. By comparing Figs. 5.3e and 5.3f with Figs. 5.3g and 5.3h, after the single PV requirement was applied, no significant reduction of the background interactions is observed, but a decrease in the outermost part of the central band is seen. It was found that events with more than one reconstructed PV in simulation were particularly rich in tracks from secondary particles originated in hadronic interactions with the detector material. Therefore, only events with one reconstructed PV are considered in the analysis.

The situation for pp collisions is different. Here, the average multiplicity is much lower so the reconstruction of a primary vertex is not as probable as for proton-lead interactions. For this reason, no requirement in the number of primary vertices is set. It is important to remember that pile-up is higher for pp events. Therefore, the probability of having more than one pp interaction is not negligible and a requirement on the number of PVs would produce a bias in the measurement. The luminous region in pp is also much spread along the z coordinate than in proton-lead, see Fig. 5.1.

The event selection criteria are summarised in Tab. 5.3.

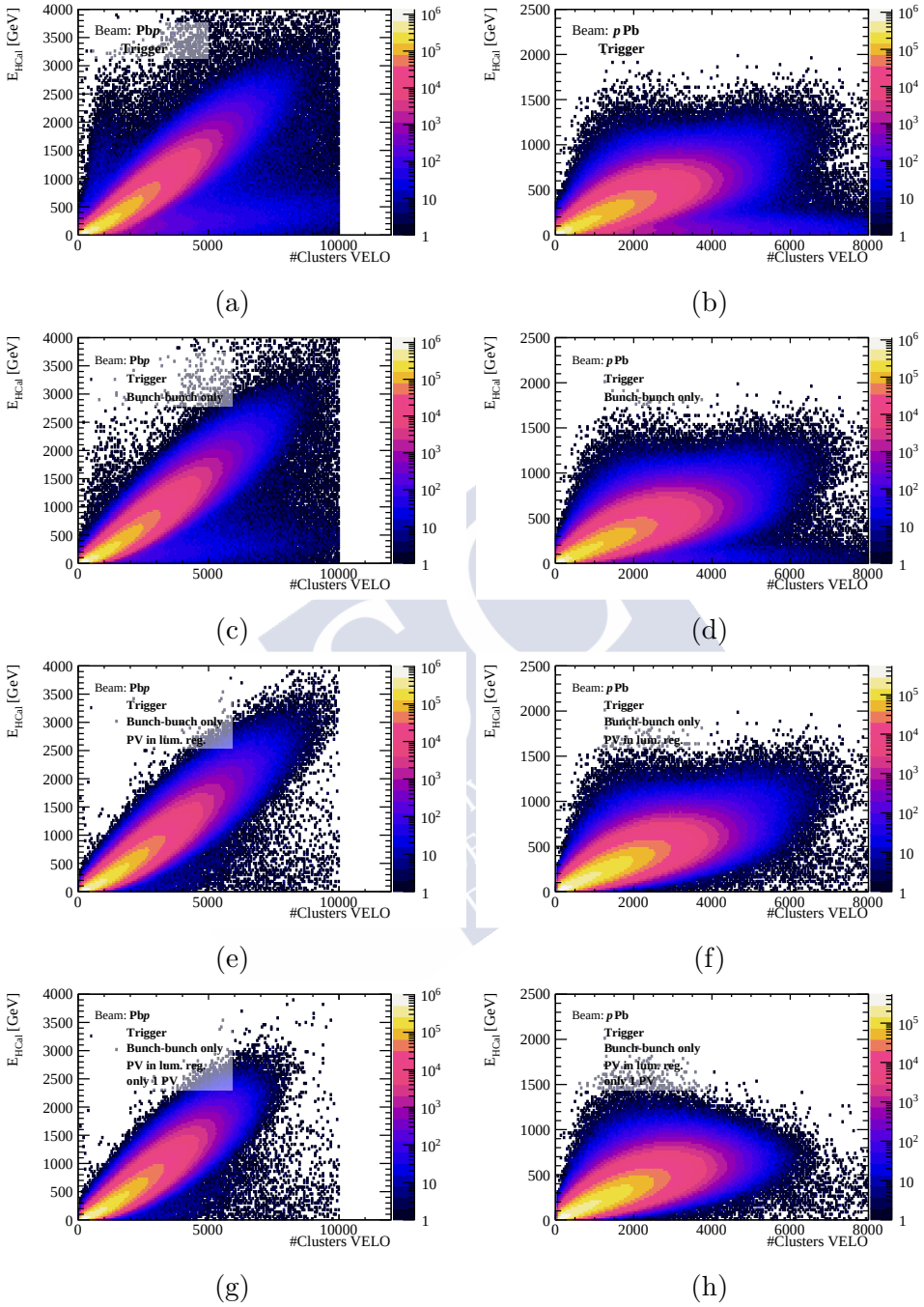


Figure 5.3: Correlation between the total energy collected in the HCAL and the number of clusters at VELO in Pbp (left) and pPb (right), at different stages of the event selection: 1) only trigger requirement (Figs. 5.3a and 5.3b); 2) adding bunch-bunch crossing (Figs. 5.3c and 5.3d); 3) adding PV in luminous region (Figs. 5.3e and 5.3f); 4) adding one PV (Figs. 5.3g and 5.3h). MD and MU datasets are merged.

5.2 Detector occupancy

Aside from kinematic variables of the prompt charged particles, (η, p_{T}), the detector response is dependent on its occupancy, defined as the particle multiplicity within the acceptance for an event. There are different ways to assess the occupancy that rely on information provided by the different LHCb subdetectors. For instance, the multiplicity in the tracking stations (VELO, TT, IT or OT) or in the calorimeters (SPD, PrS, ECAL and HCAL). Different measurements of the occupancy, which are correlated, can be considered: the number of hits in the detector, the number of clusters or the number of reconstructed tracks. As this analysis relies upon charged track reconstruction, characterising the occupancy in the tracking stations is crucial.

The detector occupancy scenarios are different in proton-lead and proton-proton collisions. In the former, a higher detector occupancy is observed due to the much higher particle multiplicity of the primary interaction. Another difference is that proton-lead collisions were recorded with very low pile-up, practically ensuring a single interaction per event, while in pp collisions the pile-up was higher, making events with multiple interactions significant. Additionally, in pp collisions the pile-up was not uniform during the data-taking, which can have an impact in the measurement.

5.2.1 Proton-lead collisions

A general concern in LHCb heavy-ion analyses is that simulation generally underestimates the occupancy of the detector. This feature mainly arises in the Pbp, but also in the pPb, configuration and has to be considered when extracting efficiencies and corrections from simulation, since they depend on the event occupancy.

The issue can be seen in Figs. 5.4 and 5.5, where different occupancy variables are shown for the pPb and Pbp configurations, respectively. The considered variables are the cluster multiplicity in the tracking detectors (VELO, Figs. 5.4a and 5.5a; TT, Figs. 5.4b and 5.5b; IT, Figs. 5.4c and 5.5c; OT, Figs. 5.4d and 5.5d) and the number of hits in the SPD (Figs. 5.5e and 5.4e). These distributions are produced after the event selection described in Sec. 5.1 with the additional requirement of at least one reconstructed long track within the LHCb detector acceptance. The high multiplicity tail underestimation in simulation (green line) is evident when compared to data (black line). Also, the shape distribution is different for both samples, specially for the number of VELO clusters. Figs. 5.4g and 5.5g show the number of selected candidates per event.

To address the data-simulation differences a weighting of simulated events is implemented. Weighted distributions of the occupancy variables in simulation should correct the tail underestimation and the shape disagreements with data. The weighting is performed with a multivariate method using the *Gradient Boost Reweighting* algorithm. The procedure is described in [156] and implemented in the `hep_ml` python package [157]. It allows to weight multidimensional distributions with correlations among the considered variables. The idea is to simultaneously adjust several occupancy variables to give a better

description of the detector occupancy.

The weighting procedure is implemented in training and testing stages. In the training stage, a multivariate classifier is fitted by comparing a simulation training sample to a target multivariate distribution: the chosen occupancy variables in data. In the testing stage, the trained classifier is used to compute the weights of a statistically independent sample of simulation and validate the resulting distributions. After that, the classifier is ready to compute weights for the entire simulation sample. No difference in the occupancy behaviour is observed between MD and MU events and samples are jointly considered. A 95% of the total simulated sample is used for training and a 5% is used for testing. This asymmetry is needed to maximise the amount of high multiplicity events in the training stage, since the main challenge for the procedure is to fix the high multiplicity tails.

In simulation, all events passing the minimum bias trigger are considered for reweighting. In the data sample, the events are required to have one reconstructed primary vertex within the luminous region. This requirement is added to avoid effects in the occupancy distributions due to background interactions, not present in the simulation. The performance of the weighting with different sets of occupancy monitoring variables was studied. Five representative sets are reported:

1: VELO, TT, IT and OT clusters (every tracking station).

2: VELO clusters and SPD hits.

3: VELO and TT clusters and number of reconstructed tracks.

9: VELO, TT, IT and OT clusters and SPD hits.

10: VELO, IT and OT clusters, number of reconstructed tracks and SPD hits.

Set 1 includes the number of clusters from every tracking station. This set is taken as a baseline. In particular, the number of VELO clusters is always considered since its distribution shape is rather different in data and simulation. The number of SPD hits is included in sets 2, 9 and 10 since the discrepancy of high multiplicity events seems to be more pronounced. Set 3 does not include the number of SPD hits nor the cluster distribution in any downstream trackers (IT, OT). Instead, the number of reconstructed tracks is considered. This variable relies upon track reconstruction, so it has been included also in set 10 to check for possible biases related to reconstruction artefacts at high multiplicity (ghost and clone tracks).

The resulting simulation weighted distributions are shown in Figs. 5.4 and 5.5, together with the data and the unweighted simulation distributions. Figs. 5.4h and 5.5h show the distribution of weights per event for each set. The shape of the different distributions of weights is similar for the different sets. Set 2 has more events with lower weights than set 3. The reason is the inclusion of the number of SPD hits in set 2, which increases the relevance of high multiplicity events while decreasing the impact of low multiplicity events. Figs. 5.4e and 5.5e show that set 3 is not optimal at reproducing the number of SPD hits distribution in data.

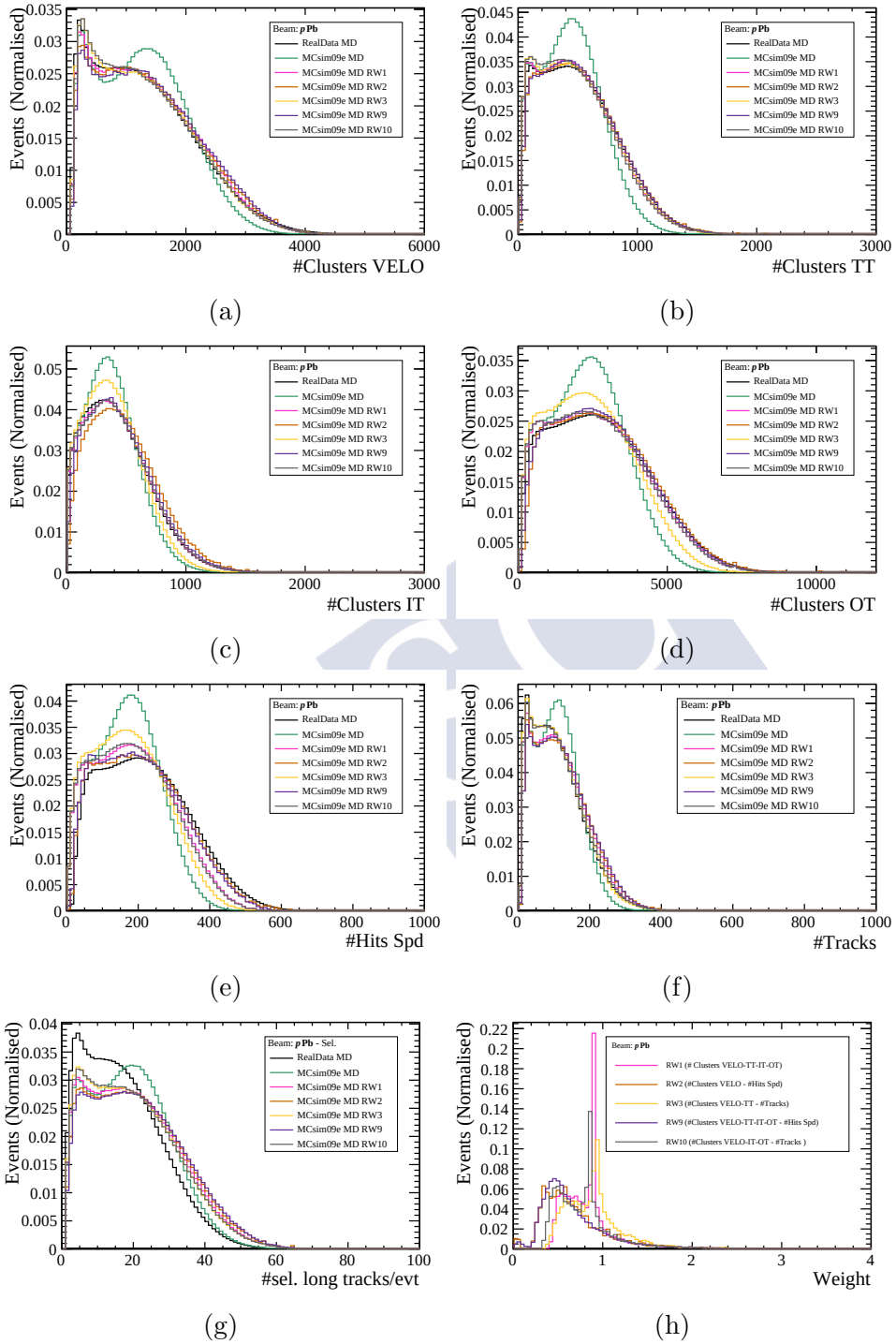


Figure 5.4: Comparison of several occupancy monitoring variables in simulation and data for the pPb MD configuration. In Figs. 5.4a − 5.4g the weighted simulation is shown using different sets of weights, labelled as RWx with x being the set number. Fig. 5.4h shows the weight distribution for each set and its legend indicates the variables considered in each weight set.

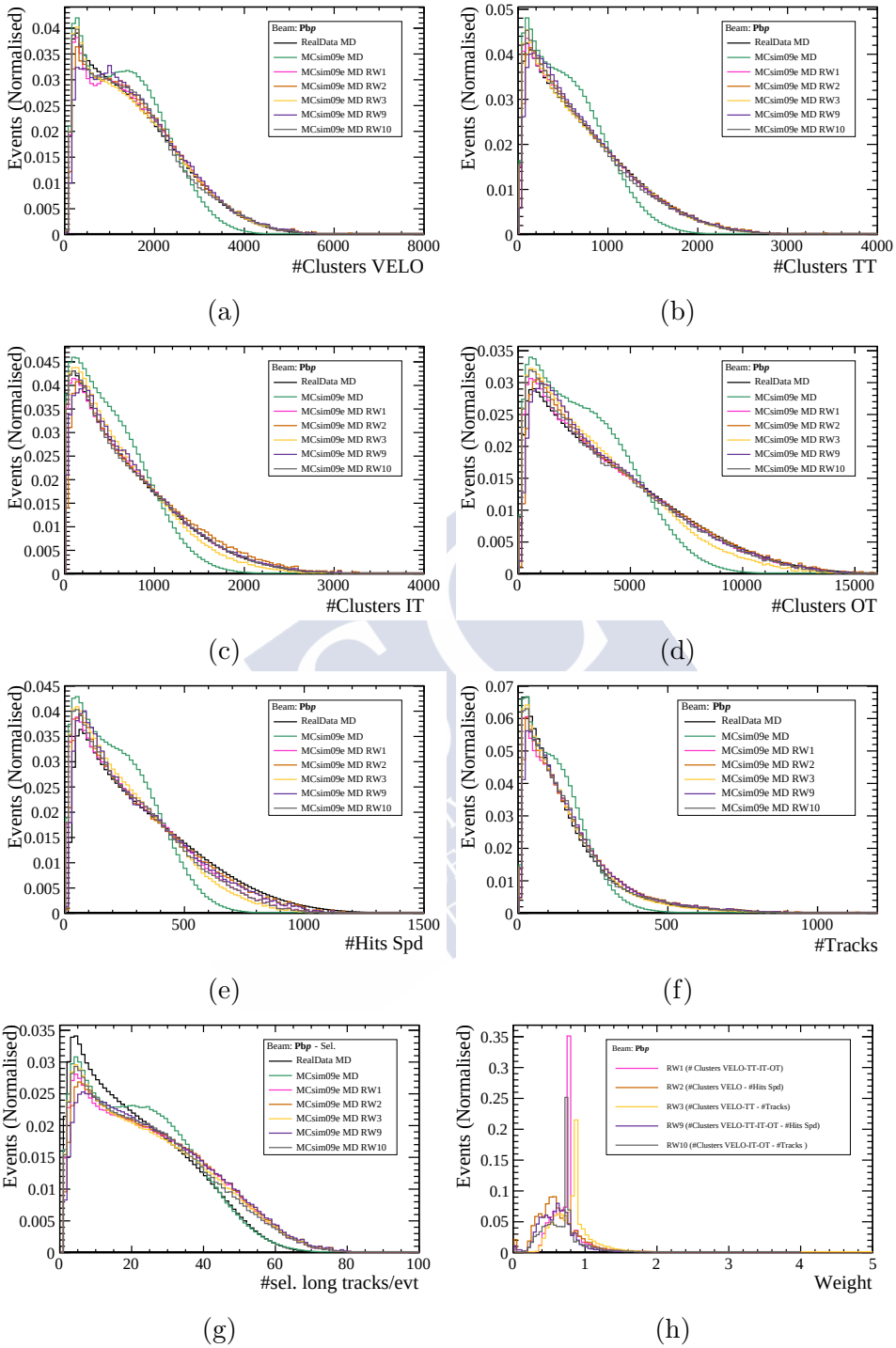


Figure 5.5: Comparison of several occupancy monitoring variables in simulation and data for the Pbp MD configuration. In Figs. 5.4a – 5.4g weighted simulation is shown using different sets of weights, labelled as RWx with x being the set number. Fig. 5.4h shows the weight distribution for each set and its legend indicates the variables considered in each weight set.

From the comparison of the different weight sets, it can be concluded that sets 1, 2, 9 and 10 are rather equivalent and give a good description of detector occupancy. For this reason, any of them can be used to reproduce the occupancy distribution in data. The comparison of the results with different sets will be useful to determine the systematic uncertainties associated to the weighting procedure (see Sec. 7.1.2). Due to the underestimation of the high multiplicity tail in the downstream detectors with set 3, this set will not be considered for systematic uncertainties evaluation. An additional cross-check is to evaluate the impact of the weights in the distributions of charged particle candidates. This is left for Sec. 5.4.2, where the candidate selection has been introduced.

5.2.2 Proton-proton collisions

Contrary to proton-lead events, in *pp* events the event occupancy in simulation does not underestimate the event multiplicity as compared to data, see Fig. 5.6. The cluster multiplicity in the VELO, TT, IT and OT are shown in Figs. 5.6a − 5.6d, the number of reconstructed primary vertices is shown in Fig. 5.6e, the number of reconstructed tracks is shown in Fig. 5.6f and the number of selected candidates is shown in Fig. 5.6g. An excess of the simulation multiplicity, as compared to data, is seen for the cluster multiplicities in the OT, while the data-simulation agreement is better in the multiplicity of the other subdetectors. From the distribution of reconstructed PVs it is clear how simulation overestimates the pile-up of data. In simulation $\nu = 1.5$ which, considering an average PV reconstruction efficiency of $\sim 70\%$ (see Fig. 5.2b), implies that $\mu \sim 1$, while during data taking $\mu \sim 0.2 - 0.6$.

To study how these data/simulation discrepancies impact the analysis, the weighting procedure applied for the proton-lead sample, explained in Sec. 5.2.1, is repeated here. However, the choice of variables is different to adapt to the particularities of the *pp* sample. Four sets of weights are computed, each considering different variables:

12: Number of reconstructed tracks and primary vertices.

13: Number of reconstructed tracks.

14: VELO clusters and number of reconstructed primary vertices.

15: VELO, TT, IT and OT clusters and number of reconstructed primary vertices.

Set 12 considers the number of reconstructed tracks, which is suggested as an optimal occupancy variable for *pp* collisions in [158], and the number of reconstructed PVs. Set 13 is used to understand the effect of including the PV multiplicity in set 12. Sets 14 and 15 consider the cluster multiplicity in the tracking stations along with the PV multiplicity. Fig. 5.6h shows the distribution of weights for the different sets.

The different sets of weights computed for *pp* data will allow to check different effects in the correction factors to the charged particle yields (see Chap. 6). By comparing set 12 and 13, the impact of reproducing the pile-up can be assessed. Set 14 allows to validate set

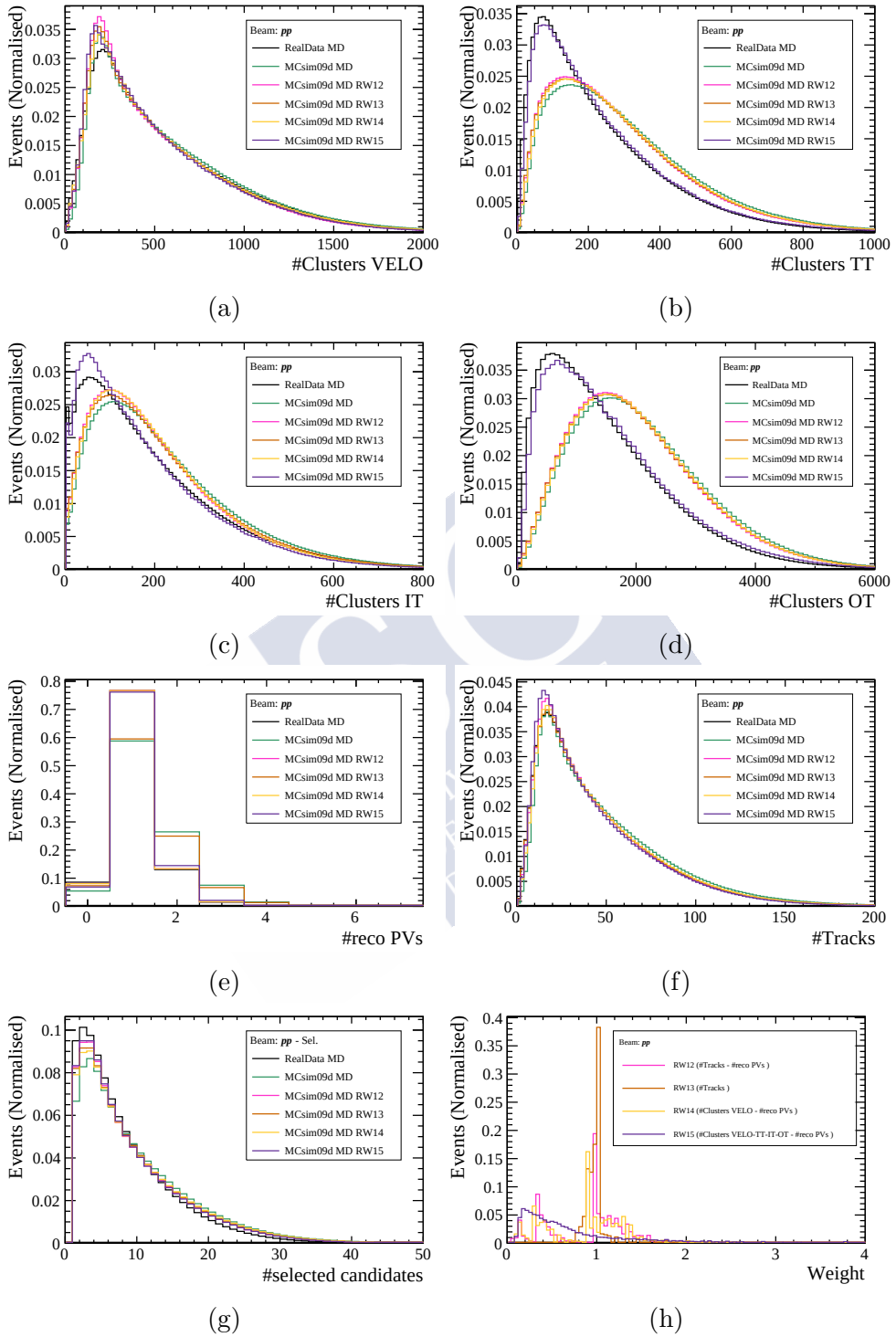


Figure 5.6: Comparison of several occupancy variables in simulation and data for the *pp* MD configuration. In Figs. 5.6a – 5.6g simulation is shown with event weighting using the different sets of weights, labelled as RWx with x being the set number. Fig. 5.6h shows the distribution of weights per event for each set.

12 by substituting the number of reconstructed tracks with another occupancy variable, the number of VELO clusters. Finally, set 15 allows to check whether it is necessary to adjust the occupancy in every tracking subdetector along with the PV multiplicity. Every set will be considered to compute the corrections and the decision on which set is more appropriate is left for the discussion of the different correction factors. Since the average event multiplicity is lower on pp, the effect of the weights on the analysis is expected to be lower than in proton-lead events.

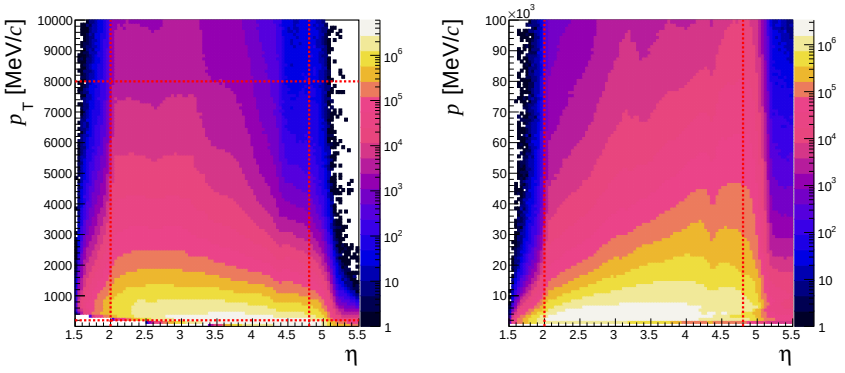
5.3 Candidate selection

Tracks reconstructed by the LHCb tracking system are used as proxies for prompt charged particles. The procedure to reconstruct tracks at LHCb has been explained in Sec. 3.2.2. The tracks for this analysis are extracted from the `TES["Rec/Track/Best"]` location in the DST files. Only long tracks are considered as *candidates* for charged particles because are optimal for physics: they have a good momentum resolution and they are likely to correspond to actual prompt charged particles. Additionally, every candidate needs to meet a series of selection requirements, related to kinematic constrains (Sec. 5.3.1), and other criteria to suppress background candidates (Sec. 5.3.3). A description of the different sources of background candidates is given in Sec. 5.3.2. Finally, a summary of the full event and candidate selection is presented in Sec. 5.3.4.

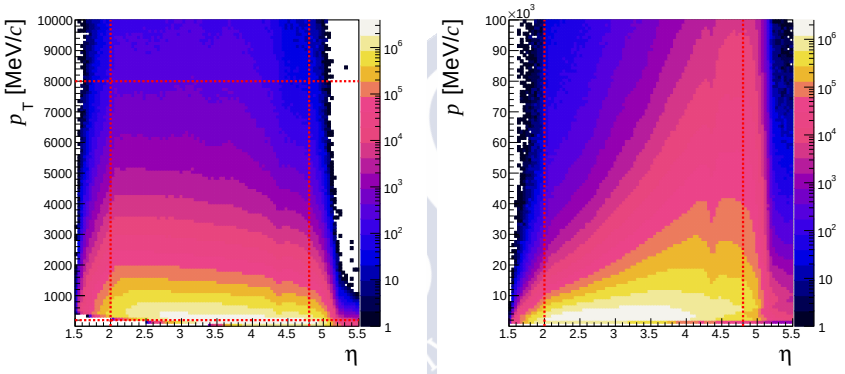
5.3.1 Kinematic range

The kinematic range of the measurement is constrained due to the acceptance of the LHCb tracking system [113]. In order to reconstruct a long track, hits from the tracking detectors before and after the magnet are required. Only particles with enough momentum travel across the magnet and leave a signature in the downstream tracking stations. For this reason, only tracks with $p > 2\,\text{GeV}/c$ are considered for the analysis. Regarding the angular coverage, the tracking system reconstructs tracks from 10 mrad to 300 mrad in the bending plane, which corresponds to $2 < \eta < 5$. It is relevant to note that most reconstructed tracks in $4.8 < \eta < 5$ do not have hits in the TT stations, resulting in a lower reconstruction efficiency and higher fake track rate in that interval (fake tracks will be discussed below in Sec. 5.3.2). As a result, the analysis is limited to the $2 < \eta < 4.8$ region. The kinematic regions of the analysis are shown in Fig. 5.7. These figures show the distribution of tracks in the data samples, directly taken from reconstruction, and after considering the fiducial acceptance requirements.

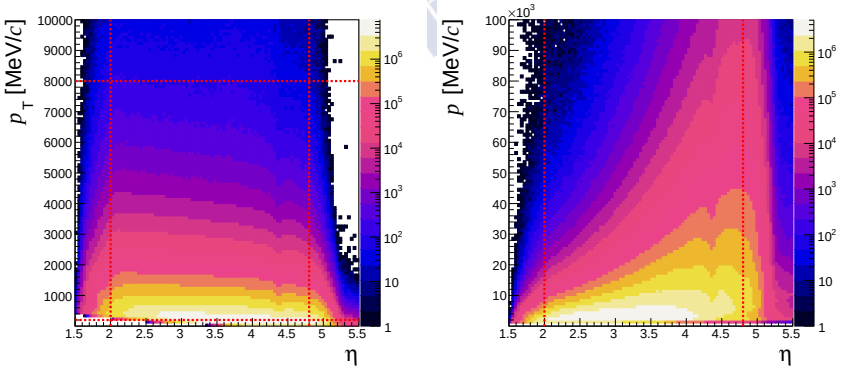
In Figs. 5.7a, 5.7c and 5.7e, the distributions of long tracks in (η, p_{T}) reconstructed by LHCb are shown for Pbp, pPb and pp configurations, respectively. The η boundaries are shown with red lines. Similar plots for (η, p) are shown in Figs. 5.7b, 5.7d and 5.7f. The low p_{T} region in the plot evidences how the momentum restriction imposes a kinematic limit for low p_{T}. The limit depends on the η value, ranging from $\sim 750\,\text{MeV}/c$ at $\eta = 2$ to



(a) Pbp MD (b) Pbp MD



(c) pPb MD (d) pPb MD



(e) pp MD (f) pp MD

Figure 5.7: Fiducial region covered in Pbp (top), pPb (middle) and pp configuration for the MD sample. The plots are made with the distribution of candidates with respect to p_{T} and η. Red lines indicate the region selected for the analysis. No difference is observed in MU with respect to MD.

$\sim 50\,\mathrm{MeV}/c$ at $\eta = 4.8$. This constraint is considered in the binning selection and will be addressed in detail in Sec. 6.1. In all figures a depression is seen around $\eta = 4.38$, which is provoked by the presence of the 25 mrad conical beam pipe inside the RICH1 [159] detector. The reconstruction efficiency is reduced there both in data and simulation. Additionally, the upper and lower limits for p_{T} are considered. A lower limit in p_{T} is set to $200\,\mathrm{MeV}/c$ since an important reduction of the reconstruction efficiency and increase of the background candidates is observed for lower p_{T} values. The upper limit is set to $p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$, due to the strong increase of the fake tracks fraction and limitations from the size of the simulated samples.

Overall, only reconstructed long tracks within the $2 < \eta < 4.8$, $p > 2\,\mathrm{GeV}/c$, $0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$ region are considered. A definition of (η, p_{T}) bins is proposed to avoid including kinematic regions out of the fiducial acceptance due to the momentum requirement (see more details in Sec. 6.1). The pseudorapidity range is indicated here in the laboratory system. It must be noted that the goal of this measurement is to calculate the nuclear modification factor to compare proton-lead with proton-proton collisions, so the relevant magnitude is η_{cms}. Due to the boost indicated in Chap. 4 its range is shifted with respect to the laboratory system to $1.6 < \eta_{\mathrm{cms}} < 4.3$ in pPb and $-5.3 < \eta_{\mathrm{cms}} < -2.5$ in Pbp, while $2 < (\eta_{\mathrm{cms}} = \eta) < 4.8$ in pp.

5.3.2 Background candidates

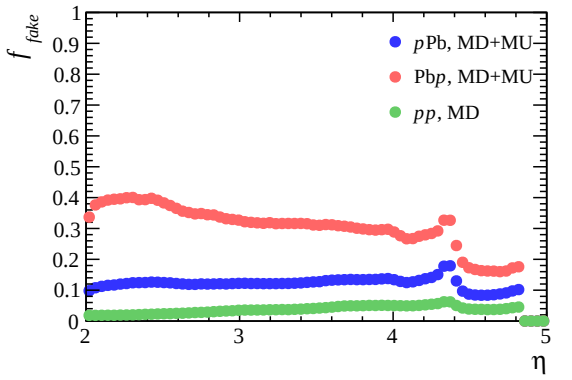
A *background candidate* is a candidate which does not correspond to a prompt charged particle. A crucial point of the analysis is to identify the different background sources and to follow specific strategies to suppress them and estimate their presence in the final sample. Three sources of background are considered: *fake tracks*, *clone tracks* and *secondary particles*. Their origin and description will be addressed in the following.

5.3.2.1 Fake tracks

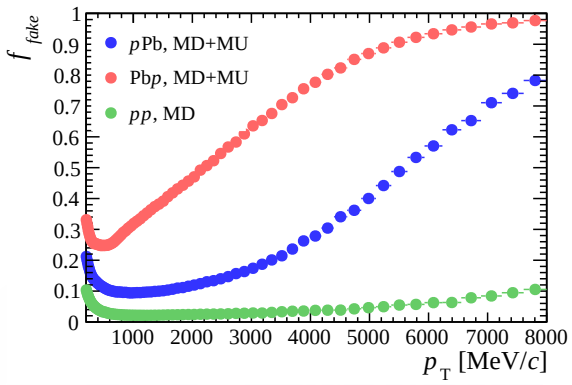
A track is classified as *fake* (or *ghost*) when it does not correspond to the trajectory of a charged particle, i.e. it is an artefact of the reconstruction. In simulation a track is considered as fake if it cannot be matched to a generated particle. The matching occurs when a reconstructed track shares at least 70% of its hits with the simulated particle.

The LHCb tracking system requires a small number of hits for a track reconstruction. This feature has a very low material budget (thus good momentum resolution) and a high efficiency as advantages, but worsens some aspects of the tracking performance, such as increasing the amount of fake tracks. This is magnified in high occupancy environments. In this way, the presence of fake tracks in proton-lead collisions, specially in Pbp, exceeds that of the standard pp collisions and may require special attention.

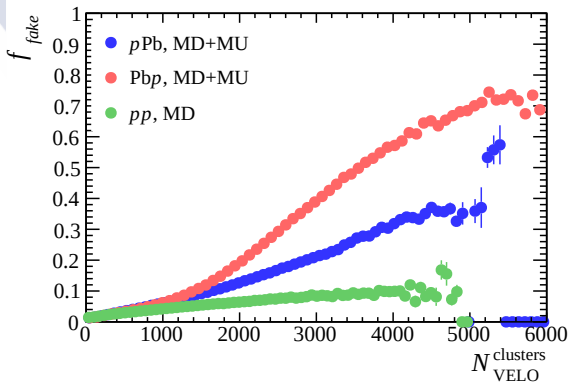
In Fig. 5.8, the ratio of non-matched tracks with respect to all tracks is shown without considering additional track quality requirements to those applied at reconstruction level. The quantity is studied with respect to η, p_{T} and the number of reconstructed VELO



(a)

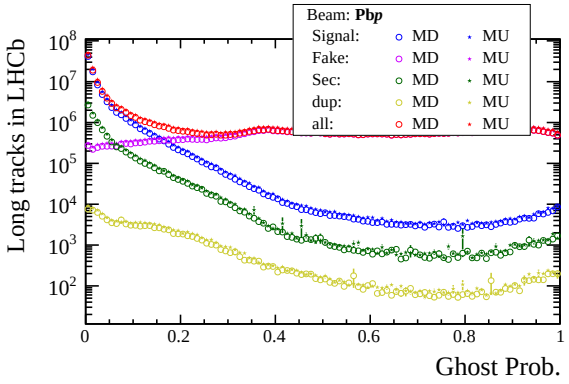


(b)

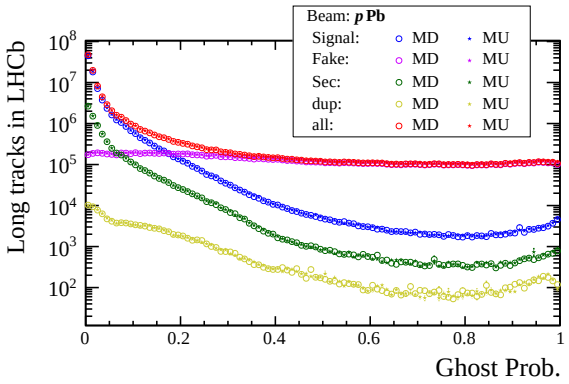


(c)

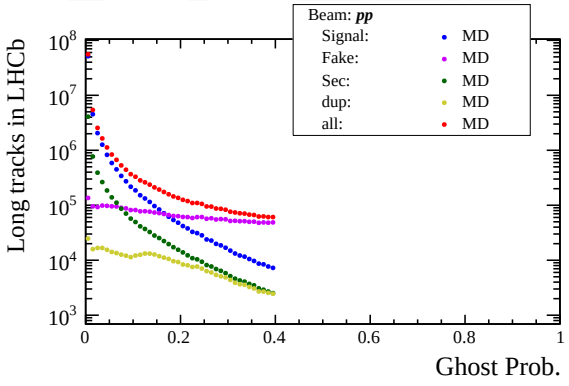
Figure 5.8: Fraction of fake tracks in simulation as a function of η (Fig. 5.8a), p_T (Fig. 5.8b) and $N^{clusters}_{VELO}$ (Fig. 5.8c), for the different configurations. The simulation is weighted with set 1 for pPb and Pbp and set 12 for pp, in order to adjust the detector occupancy to data. No track quality selection is applied apart from the requirements at the reconstruction level. Both magnet configurations are merged for proton-lead data as no difference between the two polarities is seen.



(a)



(b)



(c)

Figure 5.9: Background composition with respect to the `GhostP` variable for the Pbp (Fig. 5.9a), pPb (Fig. 5.9b) and pp (Fig. 5.9c) configurations. The simulation is weighted with set 1 for pPb and Pbp and set 12 for pp, in order to adjust the detector occupancy to real data. No track quality selection is applied apart from the requirements at reconstruction level.

clusters in the event ($N_{\mathrm{VELO}}^{\mathrm{clusters}}$). The latter characterises the global charged particle multiplicity of the event and, therefore, the detector occupancy (see Sec. 5.2 for more details). In the plots, weight set 1 for pPb/Pbp and weight set 12 for pp are considered to make the detector occupancy description similar to what is observed in data. The ratio rises with the event multiplicity and p_{T}, reaching fractions close to 90% in Pbp. The reason for the strong increase with p_{T} is that, while real track multiplicity decreases exponentially with p_{T}, fake tracks have more uniform probability to arise for any momentum value. In Fig. 5.8a, the increase at $\eta \sim 4.4$ is due to the presence of the conical beam pipe.

The amount of fake tracks in the samples is reduced by using a multivariate classifier called *ghost probability* (`GhostP`) [160]. This variable integrates information from all the tracking detectors and assigns a probability for a track to be fake. Fig. 5.9 shows the distribution of ghost probability for prompt charged particles in blue and fake tracks in magenta. The ghost probability separates very well legitimate from fake tracks.

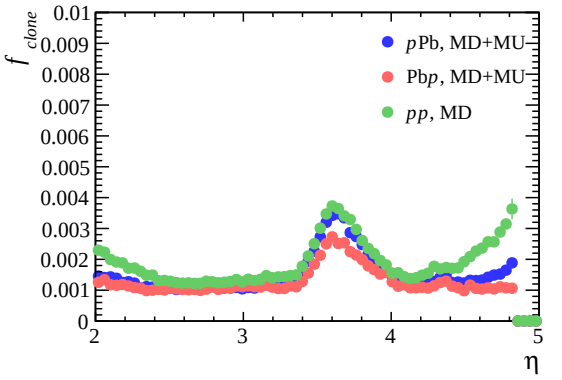
As seen in Fig. 5.9c, the reconstruction of the pp data imposes a cut in ghost probability of `GhostP`< 0.4, which is not present in proton-lead reconstruction. This is partly responsible for the large difference between pp and pPb/Pbp data in Fig. 5.8, although a lower level of fake tracks is expected in pp due to the reduced multiplicity.

In this analysis, a tighter cut in ghost probability is needed in order to reduce the fraction of fake tracks. The remaining fraction of fake tracks will be then estimated and accounted for (see Sec. 6.4 for more details). The choice of the `GhostP` requirement is addressed in detail in Sec. 5.3.3. Additionally, the matching algorithm used in reconstruction is not fully efficient, which means that some tracks marked as *ghost* in simulation might correspond to a real track. This efficiency is known as the *truth matching efficiency*, and needs to be taken into account (see for more details Sec. 6.5).

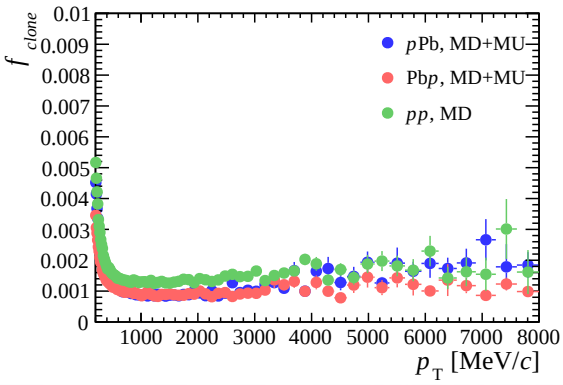
5.3.2.2 Clone tracks

Clone tracks are pairs of reconstructed tracks originated from a single real charged particle. They are reconstruction artefacts as for the case of fake tracks. A track is considered a clone if it shares at least 70% of its hits with another track. The reconstruction implemented in proton-lead and proton-proton data includes by default a requirement in the so-called *Kullback-Leibler* (KL) *clone distance* [161] to mitigate this contribution. This magnitude measures the amount of shared information between two tracks. The default cut is set to KL < 5000, which reduces the impact of clone tracks in the analysis.

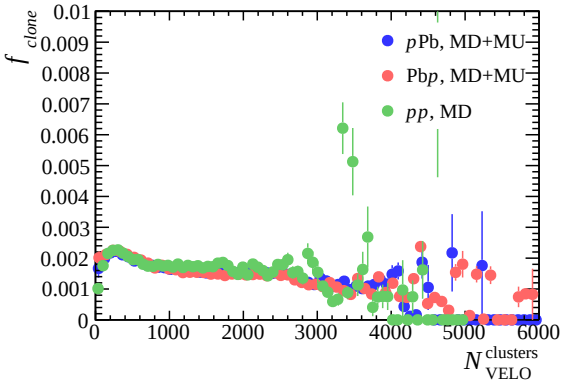
The fraction of clone tracks can be estimated in simulation from the number of reconstructed tracks that are matched to the same MC particle. When this occurs, the highest χ^2/ndf track is considered as clone. In Fig. 5.10, the fraction of clone tracks in simulation is plotted with respect to η, p_{T} and $N_{\mathrm{VELO}}^{\mathrm{clusters}}$. For all cases the fraction of clone tracks is below 0.5%. The distribution of the fraction of clone tracks peaks at low p_{T} and about $\eta = 3.6$. The multiplicity dependence is not very strong.



(a)



(b)



(c)

Figure 5.10: Fraction of clone tracks in the simulation with respect to η (Fig. 5.10a), p_T (Fig. 5.10b) and $N^{clusters}_{VELO}$ (Fig. 5.10c) for the different configurations as indicated in the legend. The simulation is weighted with set 1 for pPb and Pbp and set 12 for pp, in order to adjust the detector occupancy to real data. Data with both magnet polarities are merged for proton-lead events, since no significant differences between the two samples are observed.

5.3.2.3 Secondary particles

This background includes tracks that correspond to real particles which do not meet the prompt definition, as defined in Sec. 2.5.1. There are several sources of secondary particles, which can be classified in two groups. The first are particles originating in decays of other particles with lifetimes $> 30\,\text{ps}$, and include:

- Decays of K^0_S mesons, principally into pions.
- Decays of Λ baryons, principally into pions and protons.
- Decays of other hyperons, particularly Σ^+, Σ^-, Ξ^0, Ξ^-, Ω^-.
- Pion and kaon decays into leptons.

The second group are particles which do not originate in the primary pp or proton-lead interaction, or from subsequent particle decays, but from other interactions such as:

- Electrons from γ conversion (interaction with the detector material).
- Interactions of hadrons with the detector material producing showers.

An algorithm is dedicated to separate the prompt and non-prompt charged particles in the simulation sample according to the definition in Sec. 2.5.1. For every MC particle, a dedicated variable `IsPrompt` indicates if it is prompt. The variable is computed with an algorithm that determines if the particle is prompt by checking in an iterative way the decay chain back to the primordial pp or pPb collision. In particular, the algorithm checks three conditions:

- the particle has a mean lifetime larger than 30 ps, which includes only the particles mentioned in Sec. 2.5.1.
- none of its ancestors are particles with mean lifetime larger than 30 ps.
- the particle or any of its ancestors are not originated in a vertex type that corresponds to a interaction with the detector material (`MCVertex` :: `type` ≥ 100) [162].

If a MC particle meets these three conditions, then `IsPrompt` is set to `True`. This algorithm is applied to all particles that have been matched to a track. The secondary particle contribution is obtained by selecting tracks matched to a MC particle with `IsPrompt==False`.

The fraction of secondary tracks in simulation is shown in Fig. 5.11 with respect to p_{T}, η and $N^{\mathrm{clusters}}_{\mathrm{VELO}}$. The maximum abundance is observed at low p_{T} and then their relative impact is reduced due to the dominance of fake tracks. Their kinematic distribution is very similar to prompt particles. No important dependence with the detector occupancy is observed.

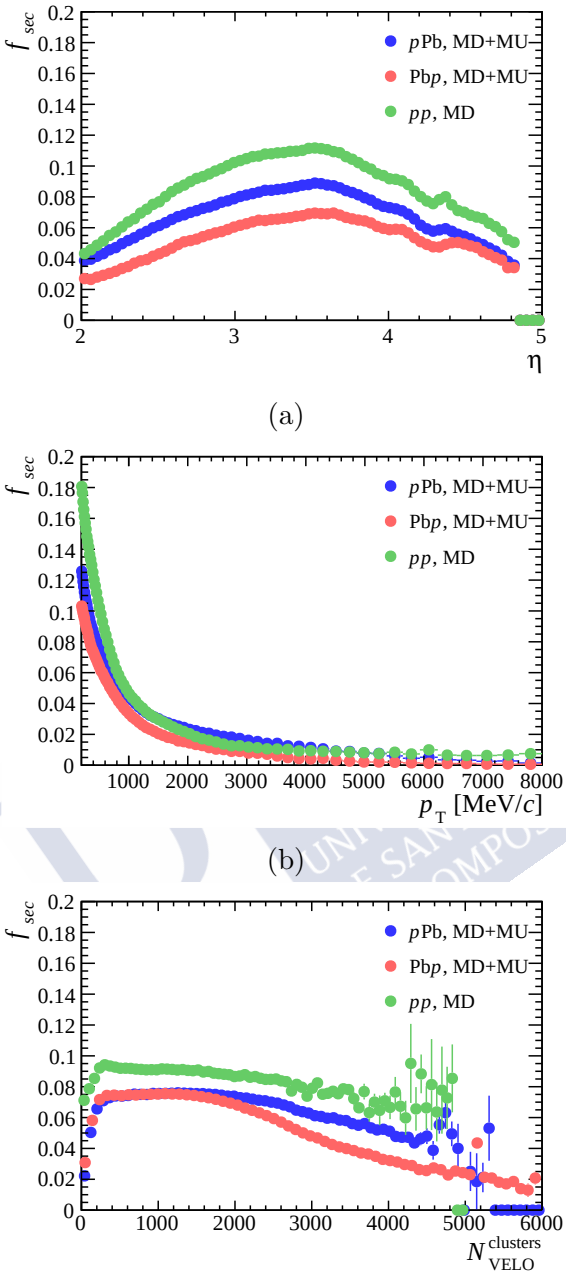


Figure 5.11: Fraction of secondary particles in simulation with respect to η (Fig. 5.11a), p_{T} (Fig. 5.11b) and $N^{\mathrm{clusters}}_{\mathrm{VELO}}$ (Fig. 5.11c) for the different configurations. The simulation is weighted with set 1 for pPb and Pbp and set 12 for pp, in order to adjust the detector occupancy to data. Data from both magnet configurations are merged for proton-lead events since no difference between the different polarities is seen.

In order to suppress secondary particles a cut in the impact parameter (IP) between the tracks and the PV is usually applied. The IP of a track is the transverse distance between the PV to the track, which is correlated to the distance that the mother particle has travelled before decaying. For this reason, it is expected for the IP to be a variable with good separation power between prompt and secondary particles. However, a difficulty arises since at least three reconstructed tracks converging to the same point are required to determine a PV, as shown in Fig. 5.2b. Thus, in the *pp* dataset not every event has a PV to compute the IP, meaning that the requirement cannot be directly applied. One solution could be excluding events without a reconstructed PV, but this could bias the measurement in *pp* towards high multiplicity events. Additionally, events with multiple PV present an additional difficulty due to the ambiguity to match reconstructed tracks with the appropriate PV where they were produced. This only affects *pp* data due to the situation explained in Sec. 5.1. The PV reconstruction efficiency is higher in proton-lead than in *pp* (see Fig. 5.2), and one reconstructed PV is already required in the event selection in proton-lead.

This issue is overcome by defining a new variable called *pseudo impact parameter* (`pseudoIP`). The variable is obtained using a PV density parametrisation as reference to estimate the origin point of every track. In particular, the PV density can be fitted to a three-dimensional Gaussian distribution:

$$\rho(\vec{r}) = \frac{1}{\sqrt{(2\pi)^3 |\det(C)|}} \exp\left[-\frac{1}{2}S^2\right] \quad \text{with} \quad S^2 = (\vec{r}-\vec{v})^T C^{-1} (\vec{r}-\vec{v}), \tag{5.2}$$

where $\vec{v}$ is the column vector that contains the average PV position $\vec{v} = (\bar{x}_{PV}, \bar{y}_{PV}, \bar{z}_{PV})^T$, $\vec{r}$ is the position and C the covariance matrix of the vertex coordinates: $C_{ii} = \sigma_i^2$, $C_{ij} = R_{ij}\sigma_i\sigma_j$. The matrix R_{ij} is the correlation coefficient between i and j and $\sigma_{i(j)}$ is the variance of the (vertex) coordinate $i(j)$. A fraction of the PV distribution sample is considered for the fits. To account for a PV, at least one long track within acceptance in the corresponding event with `GhostP` < 0.3 and $\chi^2/\text{ndof} < 3$ is required. The results of the fits for every dataset are presented in Tab. 5.4, and the projections of the gaussian in x, y and z directions are shown in Fig. 5.12. In the figure, the projections for each vertex coordinate are shown for the data sample (black), and for simulation (red or blue). The gaussian function does not perfectly describe the PV distribution in data for the x and y coordinates. However, this does not affect the performance of the `pseudoIP` estimator.

Since the transverse coordinates of the PVs are very well defined, while the longitudinal positions are smeared out (see plots in Fig. 5.1), the `pseudoIP` is defined as the transverse distance of the estimated point of origin of a track to the average PV position

$$\texttt{pseudoIP}^i = \sqrt{(x^i_{origin} - \bar{x}_{PV})^2 + (y^i_{origin} - \bar{y}_{PV})^2}. \tag{5.3}$$

Here $(x, y)^i_{origin}$ is the estimated point of origin of track i, defined as the point of a track that maximises the vertex density $\rho(\vec{r})$. This is equivalent to a maximum likelihood estimator based on the known PV density.

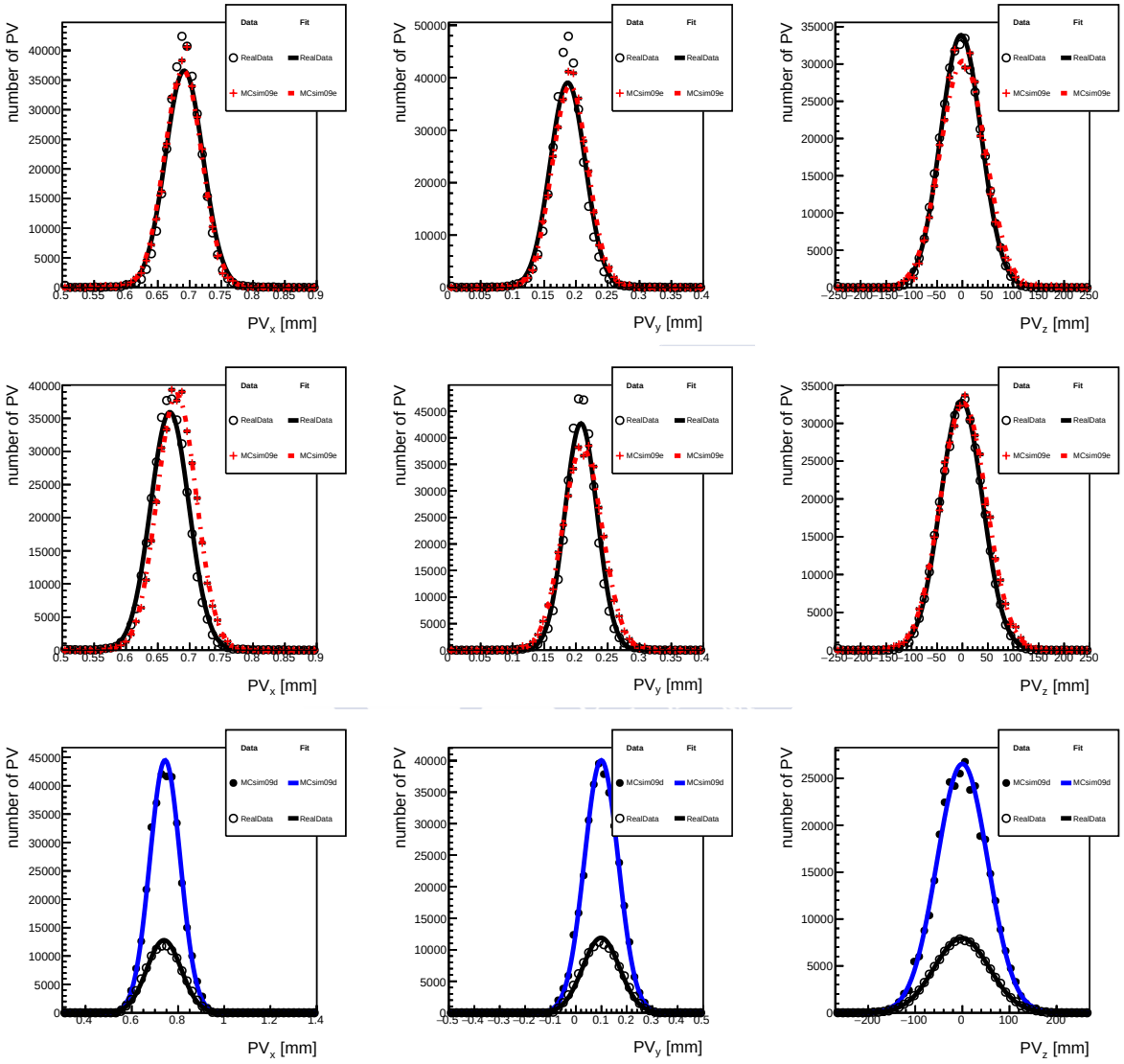


Figure 5.12: Projections of the fit to a three-dimensional gaussian function to calculate the `pseudoIP`. Configurations Pbp (top), pPb (center) and pp (bottom) in MD magnet polarity are shown. The black (red for Pbp and pPb, blue for pp) dots represent the data (simulation) and the black (red for Pbp and pPb, blue for pp) line represents the fit to data (simulation).

Close to the PV tracks can be parametrised by a straight line

$$\vec{r}(\lambda) = \vec{g} + \lambda\vec{p}, \tag{5.4}$$

where $\vec{g}$ is the column vector of the spatial coordinates of the first track state, $\vec{p}$ the column vector of the track 3-momentum components, and λ a free parameter. The point with the largest $\rho(\vec{r})$ is the one that minimises the exponential in the 3D Gaussian parametrisation S^2, such that

$$\frac{d}{d\lambda}\left([(\vec{r}-\vec{v})^T C^{-1}(\vec{r}-\vec{v})]^2]\right) = \frac{d}{d\lambda}(\vec{s}^T C^{-1}\vec{s} - 2\lambda\vec{p}^T\vec{s} + \lambda^2\vec{p}^T\vec{p}) = 0, \tag{5.5}$$

where $\vec{s} = \vec{g} - \vec{v}$. Solving this minimisation the optimal value for λ is obtained

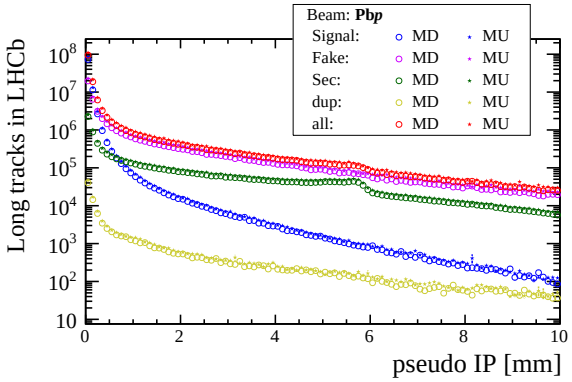
$$\lambda = \frac{\vec{p}^T C^{-1}\vec{s}}{\vec{p}^T C^{-1}\vec{p}} \text{ and } S^2_{min} = \vec{s}^T C^{-1}\vec{s} - \frac{(\vec{p}^T C^{-1}\vec{s})^2}{\vec{p}^T C^{-1}\vec{p}}. \tag{5.6}$$

Both the average position and the covariance matrix are calculated from the PV distribution in the data and simulation samples.

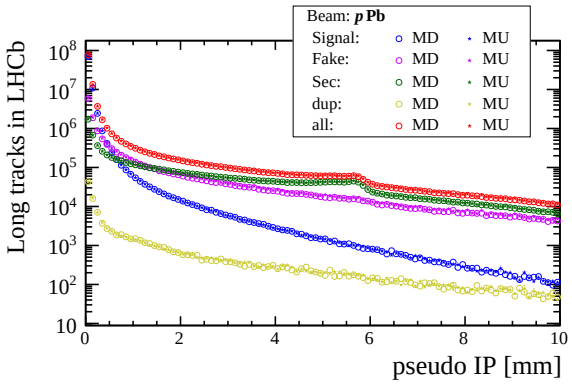
The `pseudoIP` distributions are shown in Fig. 5.13 for the different simulation samples. From the figures it can be concluded that, for every configuration, applying a `pseudoIP` threshold is a powerful strategy to remove the background of secondary particles. The `pseudoIP` is also a good variable to remove fake tracks and, since `GhostP` is the most relevant variable to discriminate against fake tracks background, a simultaneous optimisation of the `pseudoIP` and `GhostP` requirements is optimal. This is tackled in Sec. 5.3.3.

Data					
Parameter	Pbp MD	Pbp MU	pPb MD	pPb MU	pp MD
$\bar{x}_{PV}$ [mm]	0.69495 ± 0.00010	0.680725 ± 0.000095	0.67635 ± 0.00010	0.699798 ± 0.000096	0.73169 ± 0.00028
$\bar{y}_{PV}$ [mm]	0.194044 ± 0.000092	0.198792 ± 0.000095	0.212365 ± 0.000091	0.205872 ± 0.000094	0.09485 ± 0.00028
$\bar{z}_{PV}$ [mm]	-0.83 ± 0.13	0.79 ± 0.22	-2.07 ± 0.14	4.65 ± 0.23	0.58 ± 0.23
R_{xy}	-0.0343 ± 0.0035	-0.0308 ± 0.0035	0.0801 ± 0.0034	0.0512 ± 0.0035	-0.0084 ± 0.0044
R_{xz}	0.1890 ± 0.0035	-0.1497 ± 0.0034	-0.4796 ± 0.0027	-0.2290 ± 0.0034	-0.0770 ± 0.0044
R_{yz}	0.0737 ± 0.0034	0.0135 ± 0.0035	0.1090 ± 0.0034	0.0469 ± 0.0035	-0.0426 ± 0.0044
σ_x [mm]	0.029704 ± 0.000077	0.027273 ± 0.000067	0.030640 ± 0.000083	0.027324 ± 0.000070	0.06296 ± 0.00020
σ_y [mm]	0.027690 ± 0.000070	0.027548 ± 0.000071	0.030425 ± 0.000097	0.026932 ± 0.000071	0.06222 ± 0.00020
σ_z [mm]	39.609 ± 0.096	63.40 ± 0.16	42.17 ± 0.11	66.27 ± 0.17	51.24 ± 0.16
Simulation					
Parameter	Pbp MD	Pbp MU	pPb MD	pPb MU	pp MD
$\bar{x}_{PV}$ [mm]	0.689040 ± 0.000096	0.689468 ± 0.000099	0.67960 ± 0.00010	0.696607 ± 0.000090	0.74398 ± 0.00019
$\bar{y}_{PV}$ [mm]	0.190095 ± 0.000095	0.199734 ± 0.000093	0.20937 ± 0.00010	0.209268 ± 0.000091	0.09993 ± 0.00020
$\bar{z}_{PV}$ [mm]	1.69 ± 0.14	5.37 ± 0.21	0.62 ± 0.15	3.52 ± 0.20	-1.07 ± 0.16
R_{xy}	0.0039 ± 0.0033	0.0287 ± 0.0033	0.0198 ± 0.0033	0.0212 ± 0.0033	-0.0012 ± 0.0030
R_{xz}	-0.0865 ± 0.0032	0.0904 ± 0.0033	0.0066 ± 0.0033	-0.0046 ± 0.0033	0.0177 ± 0.0030
R_{yz}	0.0541 ± 0.0032	0.0289 ± 0.0033	-0.0169 ± 0.0033	-0.0011 ± 0.0033	-0.0098 ± 0.0030
σ_x [mm]	0.029475 ± 0.000068	0.030122 ± 0.000070	0.030973 ± 0.000072	0.027302 ± 0.000064	0.06574 ± 0.00014
σ_y [mm]	0.029163 ± 0.000068	0.028587 ± 0.000067	0.031044 ± 0.000072	0.027706 ± 0.000065	0.06640 ± 0.00014
σ_z [mm]	43.227 ± 0.100	62.60 ± 0.15	45.15 ± 0.10	59.90 ± 0.14	54.08 ± 0.11

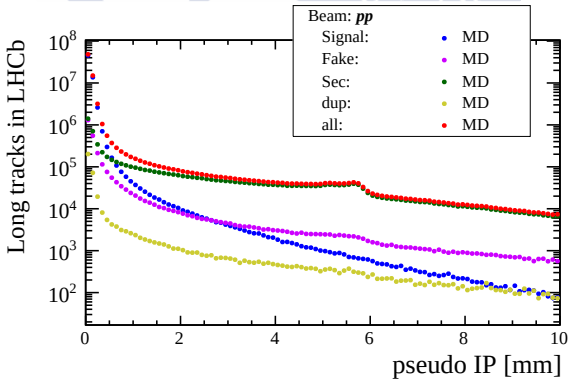
Table 5.4: Results of the parameters of the 3D gaussian fit to calculate the `pseudoIP` for the different datasets.



(a)



(b)



(c)

Figure 5.13: Background composition with respect to the `pseudoIP` variable for the Pbp (Fig. 5.13a), pPb (Fig. 5.13b) and pp (Fig. 5.13c) configurations. The simulation is weighted with set 1 for pPb and Pbp and set 12 for pp, in order to adjust the detector occupancy to data. No track quality selection is applied apart from the requirements at reconstruction level.

5.3.2.4 Background summary

The composition of the simulation sample is shown with respect to η and p_{T} in Fig. 5.14 for the Pbp, pPb and pp configurations. Both magnet polarities are displayed in the case of proton-lead and show to be equivalent with respect to the background fractions in simulation.

The kinematic distributions show how the background is dominant at the edges of the p_{T} spectra. The level of background at $p_{\mathrm{T}} > 4\,\mathrm{GeV}/c$, specially in the Pb$p$ configuration, requires the consideration of a tighter track selection, principally to remove the contribution from fake tracks. With respect to η, the background (blue) and signal distributions show a similar shape except in the vicinity of $\eta = 4.4$ where the ghost tracks fraction is enhanced. This originates in the beam pipe, as mentioned before, which worsens the reconstruction performance.

The background levels in pp before the selection are already small with respect to the signal for all the kinematic range. However, it is advisable to consider a tighter selection to further reduce the relative weight of the background in the final spectra. As it will be seen in Sec. 6.4, a precise estimation of the fraction of background in the candidate distribution of data presents difficulties, since simulation might not fully describe the background. In this way, the precision of the measurement can be optimised by minimising the background contribution.

5.3.3 Background suppression requirements

Fig. 5.14 demonstrates that a further background suppression of fake tracks in proton-lead reconstruction is desirable, specially in the Pbp configuration. It is necessary to implement additional track requirements to maximise the signal purity of the sample.

Suppression of fake tracks and secondary particles. The aforementioned `GhostP` and `pseudoIP` variables are used to optimise the signal selection by suppressing fake tracks and secondary particles. MC samples are used to find a set of selection requirements that maximises a convenient figure of merit. The goal is to find an optimal pair $(g,\, p)$ such that the selected candidates fulfill $\texttt{GhostP} < g$ and $\texttt{pseudoIP} < p$, respectively.

For that the optimisation of two different figures of merit (FoM) and the choice of the most convenient is studied in the following. The first considered FoM is the significance,

$$\mathcal{S}(g,\, p) = \frac{S(g,\, p)}{\sqrt{S(g,\, p) + B(g,\, p)}}, \tag{5.7}$$

where $S(g,\, p)$ is the amount of prompt charged particles and $B(g,\, p)$ the amount of background of fake tracks, clone tracks and secondary particles, with $\texttt{GhostP} < g$ and $\texttt{pseudoIP} < p$. The second considered FoM is the product of significance and purity, defined as

$$\mathcal{SP}(g,\, p) = \frac{S(g,\, p)}{\sqrt{S(g,\, p) + B(g,\, p)}} \cdot \frac{S(g,\, p)}{S(g,\, p) + B(g,\, p)}. \tag{5.8}$$

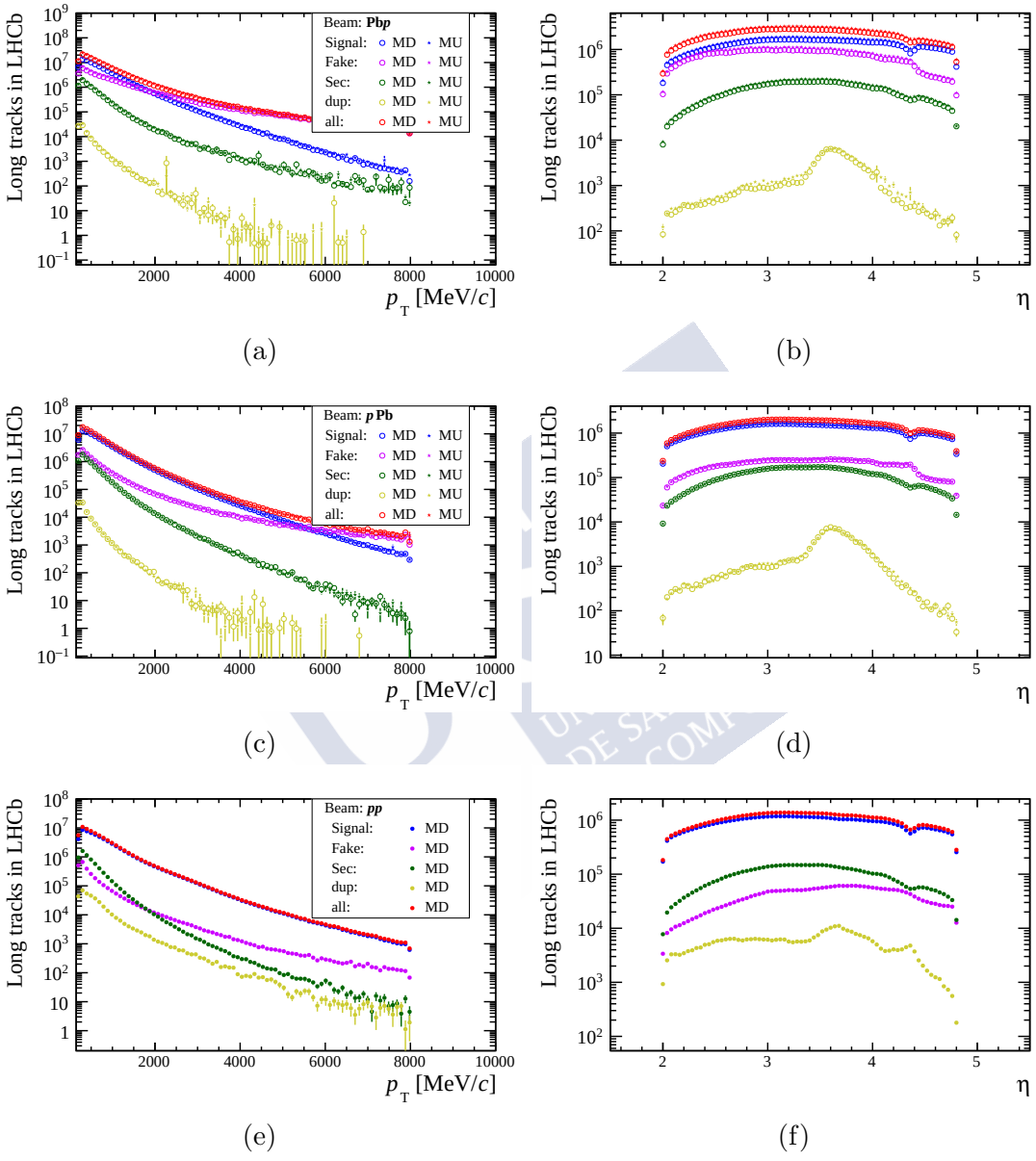
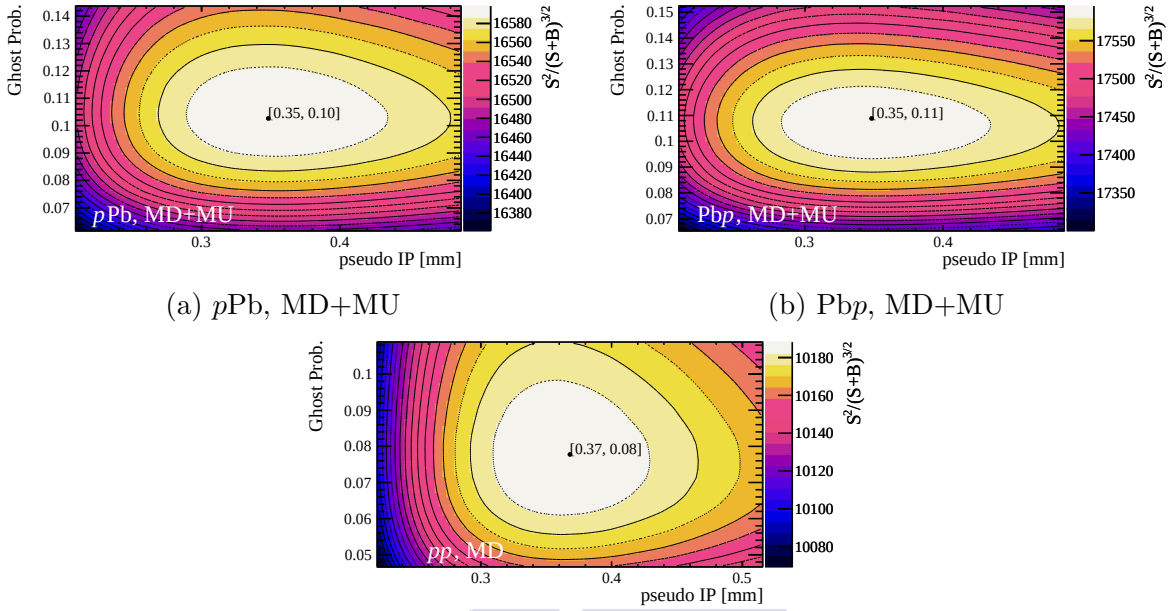


Figure 5.14: Background composition for reconstructed long tracks within LHCb acceptance with respect to p_T and η for the Pbp (top), pPb (middle) and pp (bottom) configurations.



(a) pPb, MD+MU

(b) Pbp, MD+MU

(c) pp, MD

Figure 5.15: Optimisation of the selection using $\mathcal{SP}$ (Eq. 5.8) for Pbp (Fig. 5.15a), pPb (Fig. 5.15b) and pp (Fig. 5.15c) simulated data. Weight set 1 is used in pPb/Pbp and weight set 12 is used in pp for the figure.

The result of the optimisation using the FoM of Eq. 5.8 is presented in Fig. 5.15 for each of the pPb, Pbp and pp data-sets. Since no significant difference in the values obtained for MD and MU configurations is observed, the simulation samples with different magnet polarities were combined to enlarge the sample size providing a single selection for both samples while minimising the impact of statistical fluctuations. The simulation is weighted as explained in Sec. 5.2. The different sets of weights are tested. No significant difference in the optimisation results is observed between any of the weight sets for the proton-lead datasets. A small discrepancy ($\sim 1-2\%$) is observed when comparing with non-weighted simulation. In the case of pp, there is no significant difference between any of the weight sets and the non-weighted simulation.

Tab. 5.5 shows a summary of the requirements obtained for the two presented FoMs. When evaluating the different set of requirements two considerations should be taken into account. First, the background fraction after the selection should not be too large to avoid an excessive simulation-relying background correction. Second, the selection efficiency (the fraction of signal particles kept by the requirements) should not strongly drop at any part of the (η, p_{T}) spectrum. The background rejection and the selection efficiency are correlated and their values depend on the chosen requirements. The procedures to determine the background and the selection efficiency are addressed in Secs. 6.4 and 6.3.

Due to the exponential decrease of the charged particle multiplicity with respect to

	$\mathcal{SP}$ (Eq. 5.8)		$\mathcal{S}$ (Eq. 5.7)	
	`GhostP`	`pseudoIP`	`GhostP`	`pseudoIP`
pPb	0.1026	0.34827	0.1876	0.6085
Pbp	0.1088	0.34827	0.1789	0.6085
pp	0.07775	0.36817	0.1963	0.5686

Table 5.5: Summary of the requirement sets (g, p) (`GhostP` and `pseudoIP`) obtained with the different figures of merit for each dataset.

p_{T}, low p_{T} events dominate the presented optimisation method. Additionally, background from fake tracks is not uniformly distributed along the p_{T} spectrum but concentrates in the high p_{T} and high occupancy region (Figs. 5.8b and 5.8c). As a consequence, the selection using Eq. 5.7 is not optimal to keep background levels under control in the entire Pbp data-set, that still reaches fake fractions around 50% in the high $(p_{\mathrm{T}}, N_{\mathrm{VELO}})$ region, even after considering a `GhostP` < 0.18 requirement.

The situation improves with the FoM of Eq. 5.8. By multiplying the significance by the signal purity, the selection optimisation favours a cleaner sample, and simultaneously produces a high purity of the signal and a selection efficiency above 80% in the considered (η, p_{T}) range. The possibility to further reduce the background from fake tracks with a `GhostP` requirement lower than 0.05 has also been tested. However, when doing so a strong decrease of the selection efficiency was observed at the edges of the considered η range. All in all, the requirements found with Eq. 5.8 provide the best balance between background suppression and selection efficiency and are considered for the candidate selection of this analysis.

Study of tracks sharing a VELO segment. In addition to the `GhostP` and `pseudoIP` requirements, another strategy is pursued to further reduce the background. When two different long tracks share a common VELO segment, it is likely that only one of these tracks corresponds to a prompt charged particle, being the other a background track. The background track can either be a fake or a clone track, but also a secondary particle (for instance, from the decay product of a long-lived resonance such as a K^0_{S} or a Λ). The latter occurs when a long lived particle decays after the VELO, its decay product leaves a track in the downstream trackers and, by chance, this track is matched to the VELO segment of a different charged particle.

Using the `ITrackUniqueSegmentSelector` algorithm, pairs of long tracks with a common VELO segment can be flagged. The highest χ^2/ndf track of the two is classified as *shared VELO*. This procedure is followed both in data and simulation samples.

In Fig. 5.16, the fraction of shared VELO tracks with respect to the total number of long tracks within the acceptance is shown for the data and simulation as a function of η and p_{T}. Simulated data are shown with and without applying the detector occupancy weights. The fraction of shared VELO tracks is small: $\approx 0.2\%$, $\approx 0.1\%$ and $\approx 0.06\%$ at high p_{T} for Pbp, pPb and pp events, respectively. Unweighted simulation is included in the

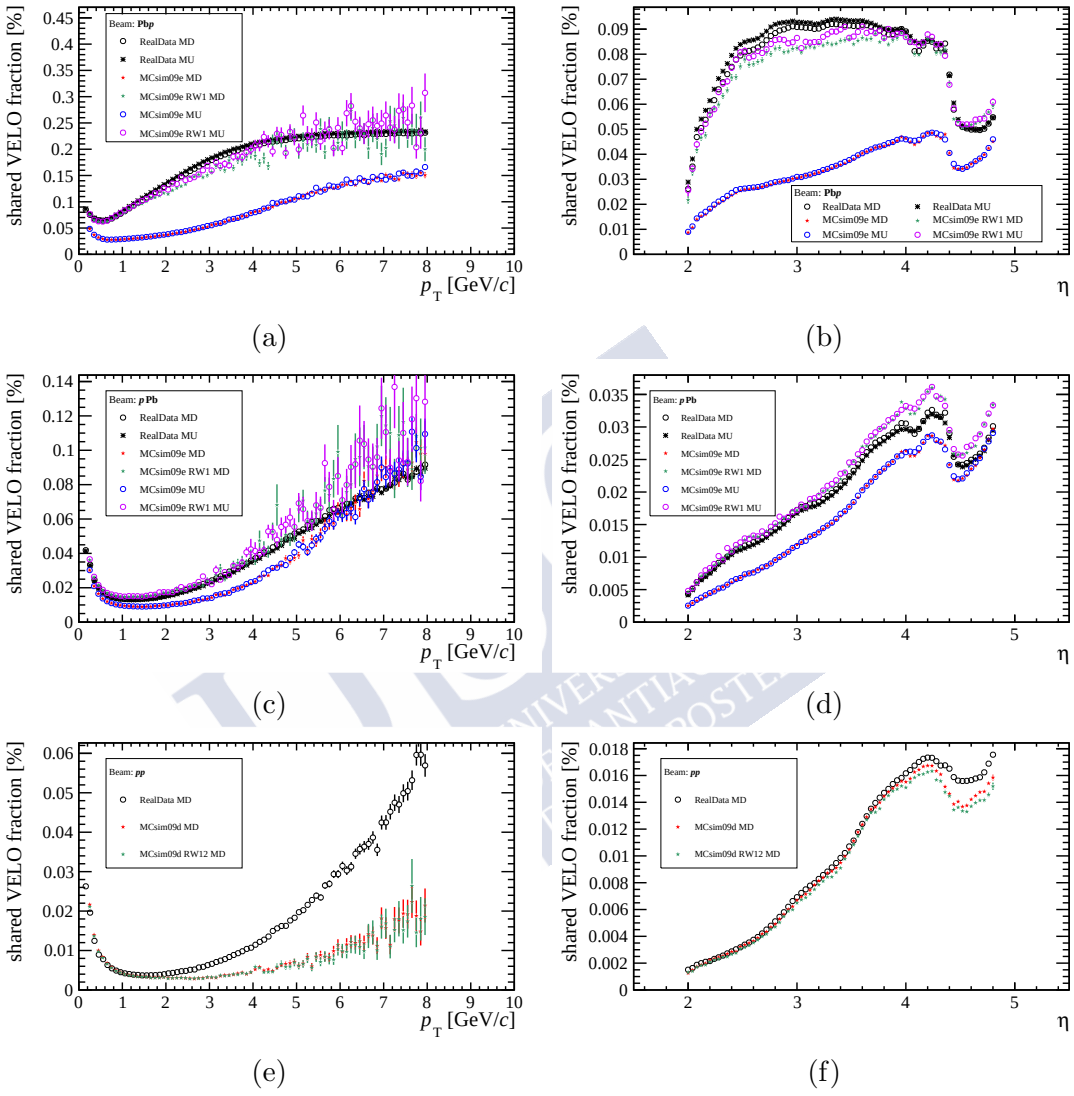


Figure 5.16: Fraction of shared VELO tracks per bin for Pbp (top), pPb (middle) and pp (bottom) configurations with respect to transverse momentum (left) and pseudorapidity (right). Simulated data are shown with and without applying the detector occupancy weights. For weighted simulation weight set 1 is considered for proton-lead events and weight set 12 for pp events.

plots to show the effect of weights at improving the data/simulation agreement, specially in Pbp. For Pbp, and to a lesser extent for pPb, the detector occupancy weights improve the data/simulation agreement. For pp data the data/simulation agreement is good for low p_{T} and throughout the η distribution. However, data show a steeper increase of the shared VELO fraction with p_{T} than simulation. The weights do not seem to significantly improve the agreement in this case.

The next step is to study the composition of the shared VELO tracks, which is shown in Fig. 5.17. From the figure, it is clear that shared VELO tracks are dominated by fake tracks, $\approx 85\%$, $\approx 78\%$ and $\approx 67\%$ in Pbp, pPb and pp, respectively. Consequently, shared VELO tracks are removed from the analysis. However these tracks are useful since they constitute a pure sample of fake tracks both in data and simulation. In fact, in Sec. 6.4 they will be used to estimate the fraction of fake tracks in data.

5.3.4 Selection summary

The selection requirements indicated in the $\mathcal{SP}$ column of Tab. 5.5 are imposed to every track candidate. Tracks flagged as shared VELO are also removed from the selection, as they were found to be mostly background. Fig. 5.18 shows the composition of the simulation samples after applying this selection. The plots display the background composition with respect to η and p_{T}, where the simulation has been weighted with set 1 for pPb/Pbp and set 12 for pp. The plot at the bottom of each figure represents the purity of the signal for every bin. In the proton-lead samples, both MD and MU configurations are displayed, showing again an excellent agreement with each other. The comparison with Fig. 5.14 demonstrates a significant reduction of the background level as a consequence of the selection, specially in the high p_{T} range.

The signal purity is above 85% for every bin. Background from secondary particles dominates at low p_{T}, mainly in pp, while fake tracks dominate the high p_{T} background for every dataset. It is worth noting the larger error bars in the purity plot for high p_{T}, specially in the Pbp sample. The reason for this behaviour are the large values for the simulation weights in that p_{T} region (Fig. 5.5h) originated in the poor statistics of the simulation sample in that range. This adds up to the lower statistics of data at high p_{T}. A larger binning in this region will be used to mitigate this issue (Sec. 6.1). However, it is the lack of simulation candidates in this regime that has prevented to cover a larger p_{T} range in this analysis. Tab. 5.6 summarises all the selection criteria for events and candidates.

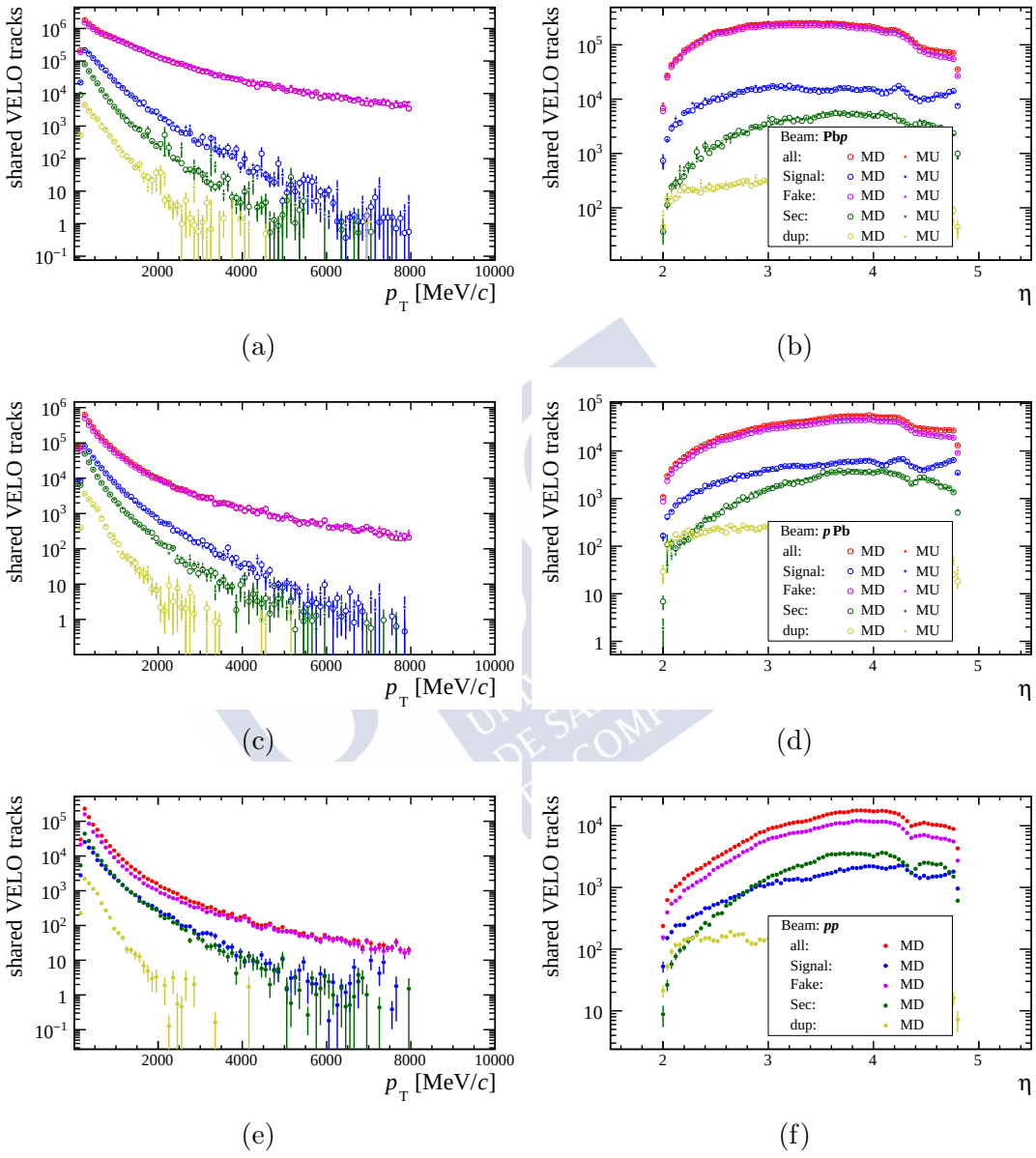


Figure 5.17: Composition of shared VELO tracks in simulation with respect to η (right) and p_{T} (left) for Pbp (top), pPb (middle) and pp (bottom) configurations. Weight set 1 is applied to pPb/Pbp and weight set 12 is applied to pp.

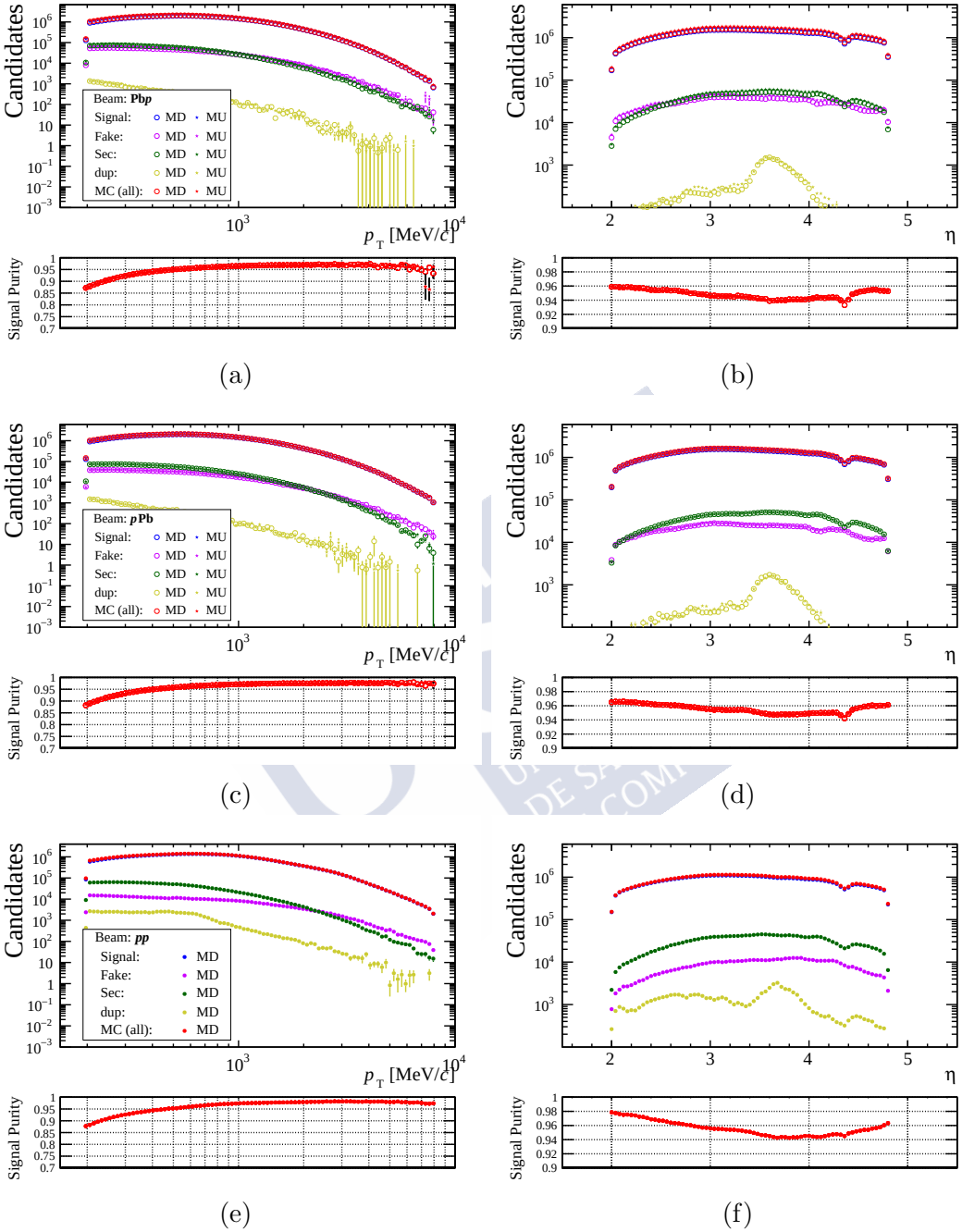


Figure 5.18: Background composition and signal purity of the distribution of candidates with respect to p_{T} (left) and η (right) in simulation for the Pbp (top), pPb (middle) and pp (bottom) dataset. Weight set 1 is applied to pPb/Pbp and weight set 12 is applied to pp.

		pp	pPb	Pbp
Event selection	Trigger	`Hlt1NoBiasLeadingCrossing`	`Hlt1MBMicroBiasVelo`	`Hlt1MBMicroBiasVelo`
	Bunch crossing type	bunch-bunch	bunch-bunch	bunch-bunch
	number of PV	-	1	1
	PV position	-	within luminous region	within luminous region
Candidate selection	Track location	`TES["Rec/Track/Best"]`	`TES["Rec/Track/Best"]`	`TES["Rec/Track/Best"]`
	Track type	Long	Long	Long
	η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
	p	$p > 2\,\text{GeV}/c$	$p > 2\,\text{GeV}/c$	$p > 2\,\text{GeV}/c$
	p_T	$0.200 < p_\text{T} < 8\,\text{GeV}/c$	$0.200 < p_\text{T} < 8\,\text{GeV}/c$	$0.200 < p_\text{T} < 8\,\text{GeV}/c$
	`pseudoIP`	0.368 mm	0.348 mm	0.348 mm
	`GhostP`	0.078	0.103	0.109
	is shared VELO	`False`	`False`	`False`

Table 5.6: Candidate selection considered for the analysis. The event selection from Tab. 5.3 is repeated here for consistency.

5.4 Validation of simulation samples

In this section, the agreement between the different data and simulated samples is studied. First, kinematic distributions of selected candidates are compared. Then, the different sets of weights obtained in Sec. 5.2 are validated by comparing the different weighted distributions with data.

5.4.1 Validation of candidate distributions

Fig. 5.19 shows the distributions of candidates in the pp dataset. Simulation is presented with weight set 12 and without any weights. None of the presented distributions are affected by the weighting. There are discrepancies at high p_{T} and low p between data and simulation. Regarding η, simulation overestimates the number of tracks for $\eta \lessapprox 3.3$ and underestimates it for $\eta \gtrapprox 3.3$. The discrepancies in the spectrum are present before and after considering the selection, so they cannot be attributed to differences in the background between data and simulation. The discrepancies probably originate from Pythia not reproducing correctly the prompt charged particle spectra. With respect to ϕ, the azimuthal angle, the simulation reproduces well the distribution from data.

The discrepancy in the η and p_{T} is studied in detail in Fig. B.2 for specific (η, p_{T})

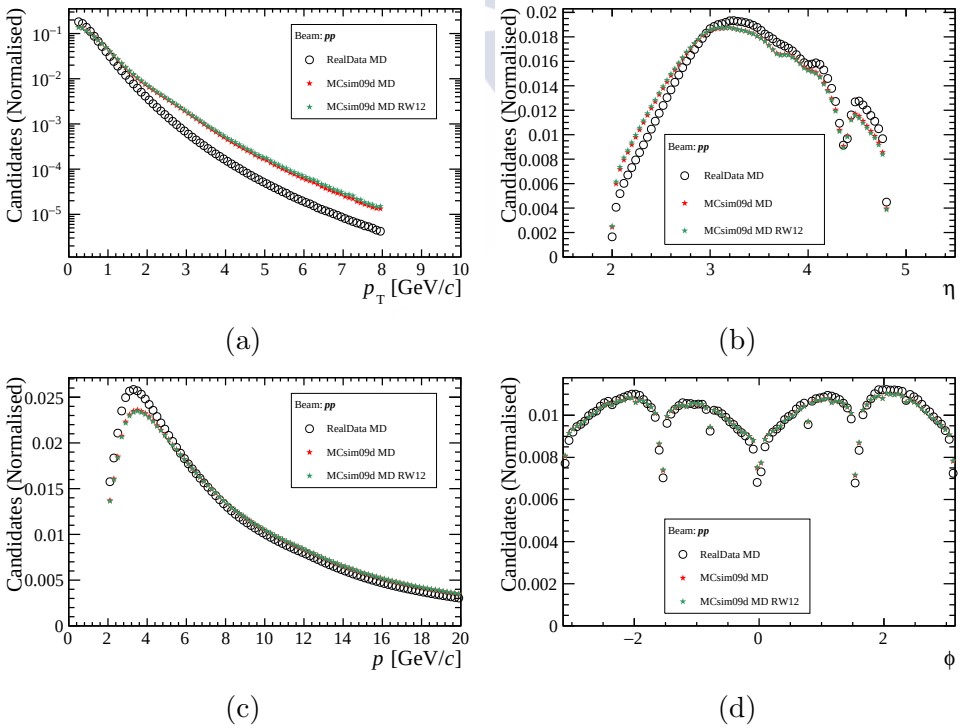


Figure 5.19: Distributions of candidate tracks with respect to p_{T}, η, p and ϕ in data and simulation for the pp dataset.

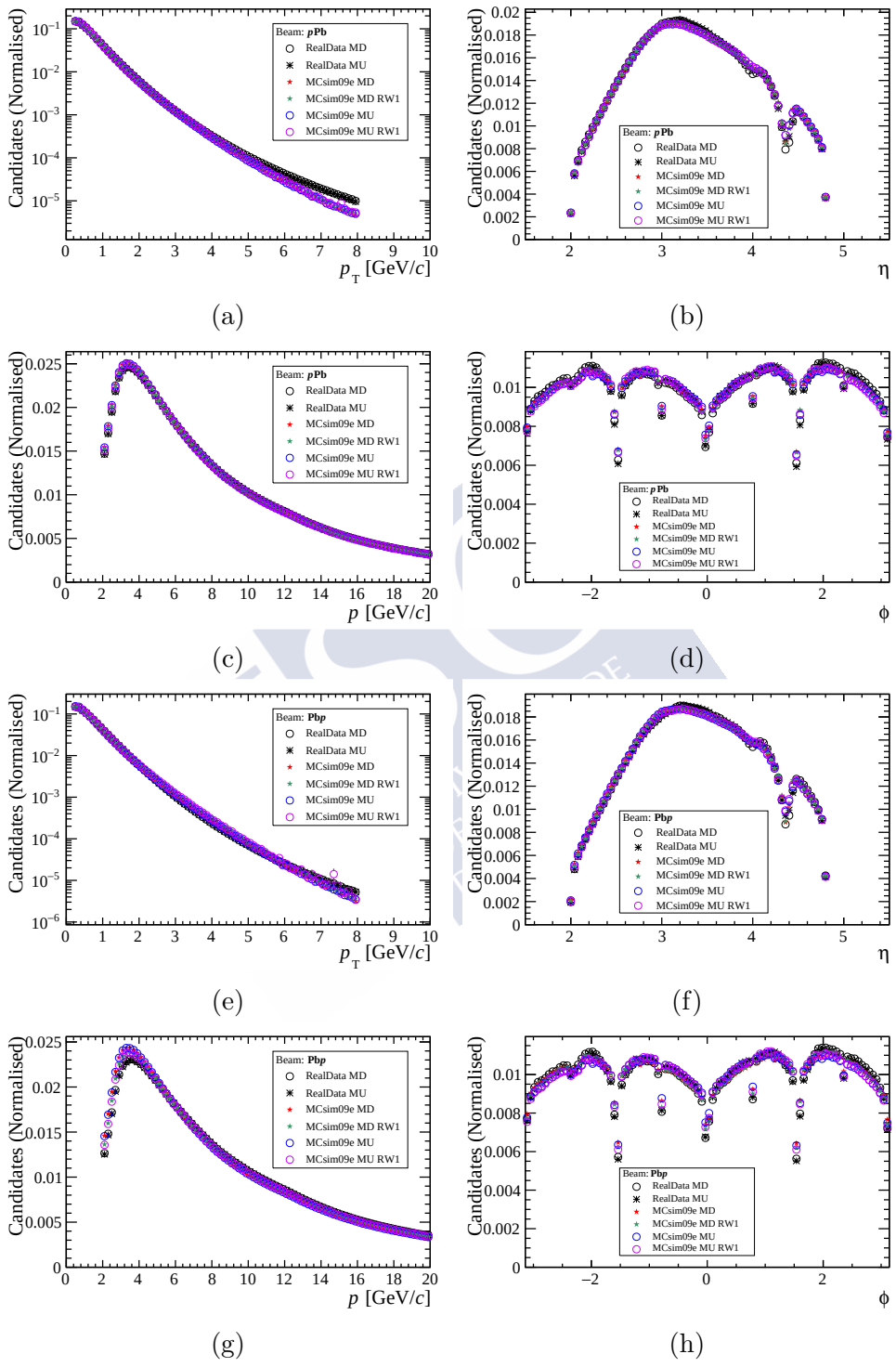


Figure 5.20: Distribution of candidate tracks with respect to p_{T}, η, p and ϕ in simulation and data for the pPb and Pbp datasets.

bins that will be used in the measurement (see Sec. 6.1 for more details). A very good agreement is seen in all the presented bins.

Fig. 5.20 presents the distribution of candidates in the pPb and Pbp datasets. The different kinematic variables show excellent agreement for both datasets. The agreement is better than for pp, which can be explained in the EPOS generator reproducing better the minimum bias production of charged particles than the Pythia generator. Weights do not significantly modify the kinematic distributions of the candidates.

5.4.2 Validation of the simulation weights

The study performed in Sec. 5.4.1 can be repeated but focusing in comparing the effect that the different sets of weights that were obtained in Sec. 5.2 have in the distributions of candidates and also in the distributions of all long tracks. The reason to study also long tracks that do not pass the selection is that they allow to see how background is affected by the occupancy weights.

In Fig. 5.21, the effect of the simulation weights in the distribution of relevant variables is checked for the proton-lead samples. This includes the p_{T} (Figs. 5.21a and 5.21b) and η of candidates (Figs. 5.21c and 5.21d), the `GhostP` (Figs. 5.21e and 5.21f) and `pseudoIP` of long tracks in the acceptance (Figs. 5.21g and 5.21h). The selection requirements are not applied to the `GhostP` and `pseudoIP` distributions to evaluate the effect of the background contribution in data and simulation. First, it is clear that applying any set of weights does not affect the p_{T} and η distributions of the candidates. Second, in the `GhostP` distribution, the discrepancy at high values is corrected by the weighting. The reason is that the number of fake tracks, which dominate in this region, is highly correlated with the occupancy. This is relevant for the background determination in Sec. 6.4, as tracks from this `GhostP` range will be used. The weights have a lesser effect in the `pseudoIP` distribution. An increase is observed in the Pbp distribution for high `pseudoIP` which is also due to the contribution of fake tracks. Other variables, such as p, ϕ or the track χ^2/ndof have been checked after including the weights. All these distributions show variations at the level of the expected increase of the fraction of fake tracks.

It is worth noting that the `pseudoIP` distributions for pPb and Pbp (Figs. 5.21g and 5.21h) are not perfectly reproduced even after the reweighting. To investigate if this is due to a bad modeling of the `pseudoIP` in the simulation, the distributions can be compared for specific (η, p_{T}) bins, as was done before for pp in Sec. 5.4.1. This is presented in Fig. B.4 and Fig. B.3 of the appendix B. The agreement for specific bins becomes much better.

For proton-proton events the same simulation weights validations is performed in Fig. 5.22. The p_{T} and η candidate distributions remain unaffected after applying any of the sets of weights. The `GhostP` and `pseudoIP` long track distributions also remain unchanged, which indicates that the background is not affected by the weights. This is reasonable since the occupancy distributions in pp simulation are already similar to data before weighting.

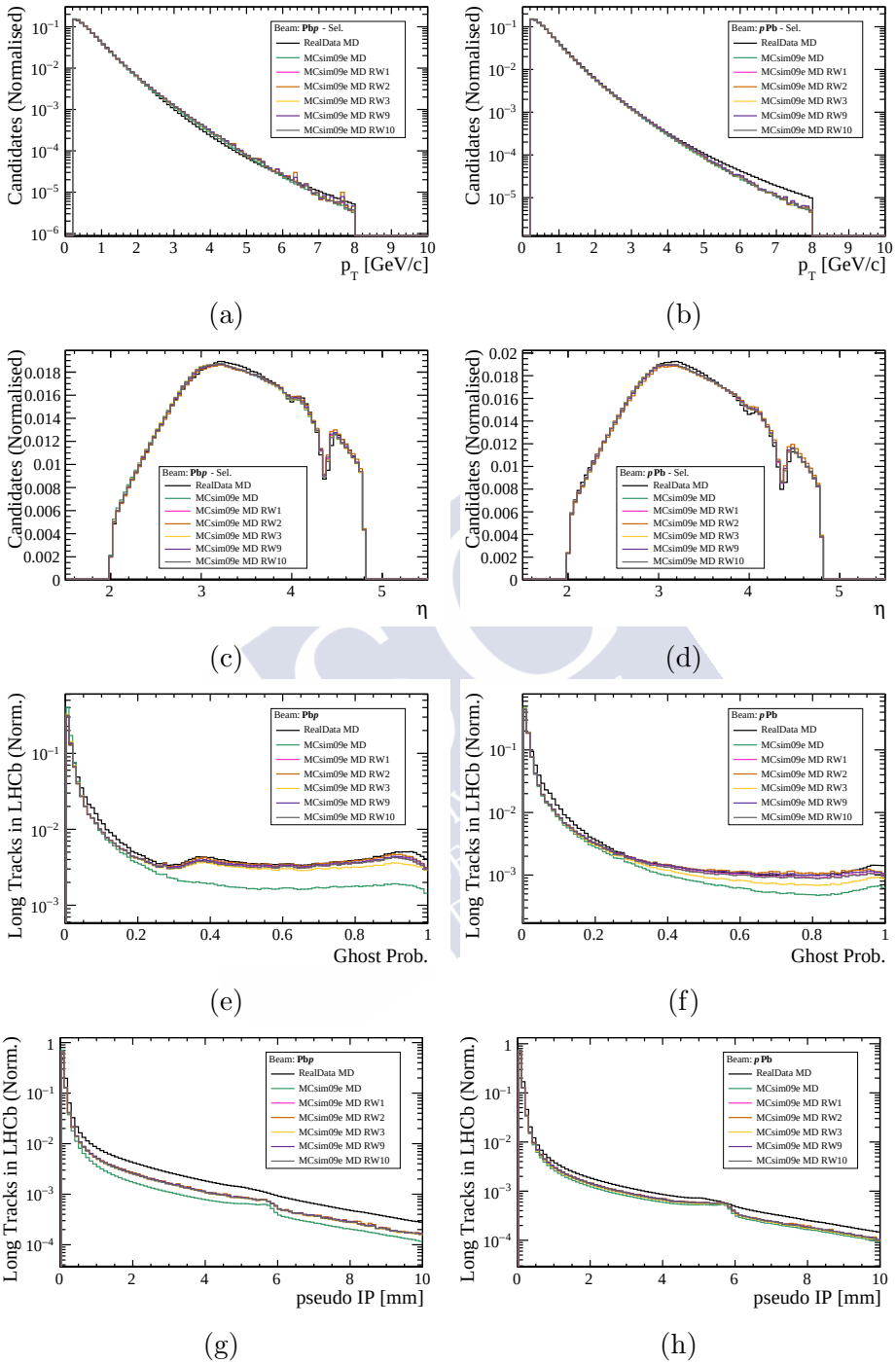


Figure 5.21: Comparison of the p_{T} and η distributions (for candidates), and `GhostP` and `pseudoIP` distributions (for long tracks within the LHCb detector acceptance) in data and simulation with different sets of weights in the Pbp (left) and pPb (right) MD configurations.

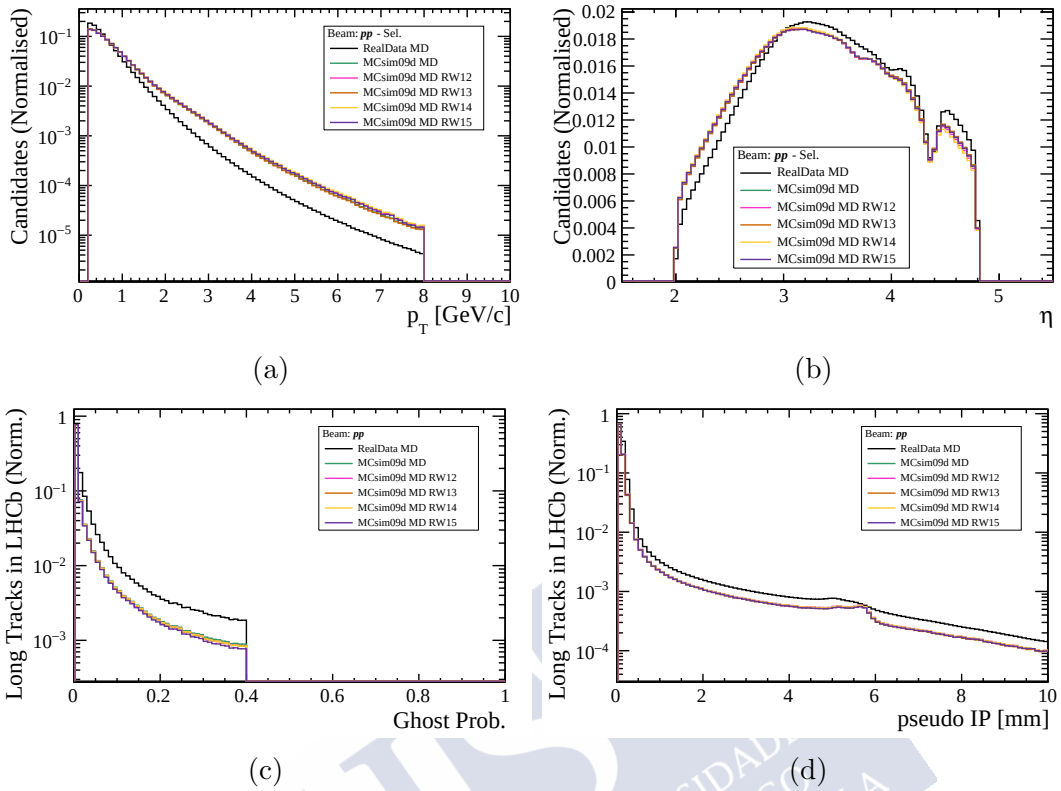


Figure 5.22: Comparison of the p_{T} and η distributions (for candidates), and `GhostP` and `pseudoIP` distributions (for long tracks within the LHCb detector acceptance) in data and simulation with different sets of weights in the pp configuration.

There is a notable discrepancy data over simulation in the `GhostP` distribution for Fig. 5.22c. As was done before, this discrepancy is studied further in the appendix in Fig. B.1, where the `GhostP` distribution is presented for specific (η, p_{T}) bins. The agreement for low values of `GhostP` is much better for all bins. An important disagreement for high p_{T} in high values of `GhostP` is seen, specially in the highest p_{T} (Fig. B.1d), which reminds to the `GhostP` distribution for data in Pbp from Fig. 5.21e. This discrepancy is due to a higher fraction of fake tracks in the data than in the simulation, and will be investigated in detail in Sec. 6.4.1.

5.4.3 Validation of the particle composition of the simulation.

The composition of the generated prompt charged particles in the simulation samples needs to be validated with data. The reconstruction efficiency (see Sec. 6.2), can be different for the different particle species included in the prompt charged particle definition of Sec. 2.5.1. Also, the validation of the abundances of K^0_{S} and Λ is useful to determine if the background from secondary particles coming from decays of these resonances is well reproduced in simulation. Results from the ALICE collaboration will be used since their

measurements of all particle species of interest are available for the considered p_{T} range, both in pp and pPb [93].

Proton-proton validation. In the measurement by ALICE, an extrapolation of the invariant yields to 5.02 TeV is made in the same study using the measurement at 2.76 TeV. These data can be used to construct the K/π and p/π ratios at 5.02 TeV in the kinematic acceptance of $-0.8 < \eta < 0.8$. Once these ratios have been obtained, they can be compared with the LHCb simulation in the same acceptance. Additionally, the ALICE collaboration has measured the invariant yield of Λ, K^0_{S}, Ξ and Ω in pp collisions at 7 TeV [163], and these data can be combined with the π measurement to construct the hyperons/π ratio. To adapt the π yield binning, an interpolation is made using the ROOT `TSpline3` class [164]. No data are available for the $\Sigma^{\pm}$ production, however the yield can be estimated with the Λ yield multiplied by the $\Sigma^{\pm}/\Lambda$ ratio in simulation, as done in [41].

The hadron/π ratios from ALICE measurements are presented in Figs. 5.23 and 5.24. In Fig. 5.23, the ALICE measurement is compared with the ratio in official pp simulation, generated with PYTHIA. In Fig. 5.24, the measurement is compared with a private sample of pp collisions generated with EPOS. In both cases, a band of $\pm 30\%$ around the central value of the fraction in simulation is considered to estimate the systematic uncertainty (see Sec. 7.1.3). As seen in Fig. 5.23, PYTHIA is unsuccessful to reproduce the hyperon/π ratio, and also the K/π and K^0_{S}/π ratios at $p_{\mathrm{T}} > 2$ GeV. The fact of PYTHIA not reproducing strangeness production in pp is already known [163, 165]. The EPOS simulation does reproduce the measured ratios within the considered uncertainty band.

LHCb has also measured ratios of prompt π^-, K^- and p at 0.9 TeV and 7 TeV [166]. Comparisons of these measurements with PYTHIA and EPOS predictions are shown in Fig. 5.25, along with a band of $\pm 30\%$ around the simulation predictions. The comparison is more limited than with ALICE data since only three p_{T} bins are available, but the acceptance corresponds to the η range of interest for this measurement. The K/π ratio at $p_{\mathrm{T}} > 1.6\,\mathrm{GeV}/c$ is well reproduced by EPOS and underestimated by PYTHIA, as for ALICE data in central rapidity.

Due to the limited p_{T} range of the measurement and the lack of information about hyperons/π ratio ALICE measurements are used to cross-check the simulation, assuming that the hadron/π ratios are independent on η in the $-5 < \eta < 5$ range. The following arguments support this assumption:

1. Neither PYTHIA nor EPOS generators predict a strong η dependence in the relative abundances, both in pp, pPb and Pbp collisions. This can be seen in Fig. 2.10.

2. The LHCb measurement of the K/π and p/π ratios [166] does not hint any strong η dependence in $2.5 < \eta < 4.5$.

In addition to EPOS and standard PYTHIA samples, an additional sample generated with a PYTHIA tune that considers string–string interactions to model collective effects [167] is studied. The sample accounts for 10^6 events that were generated using

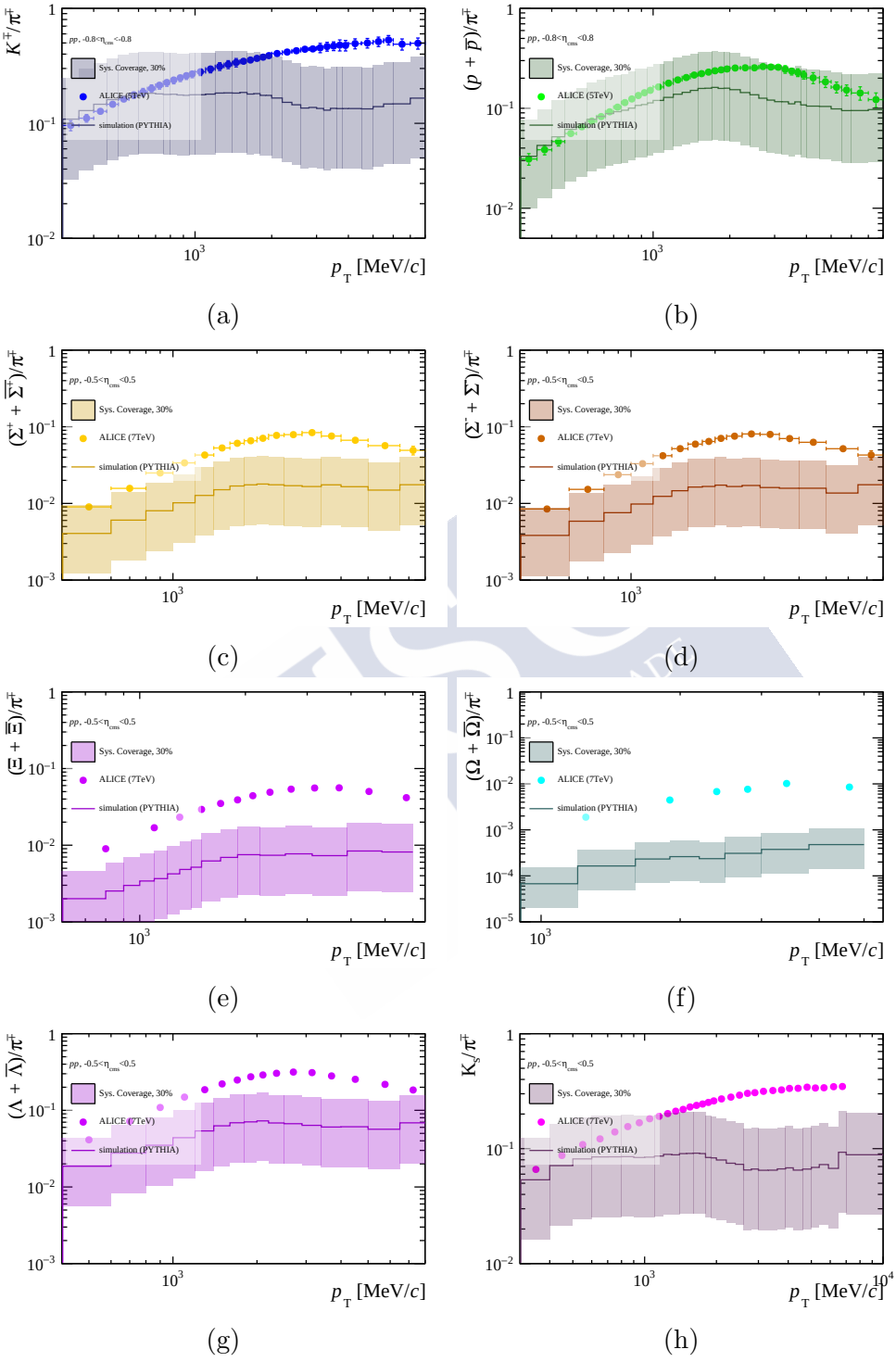


Figure 5.23: Comparison of the hadron/π ratios for pp Pythia simulation and ALICE data [93, 163]. The band corresponds to a 30% variation of the ratio.

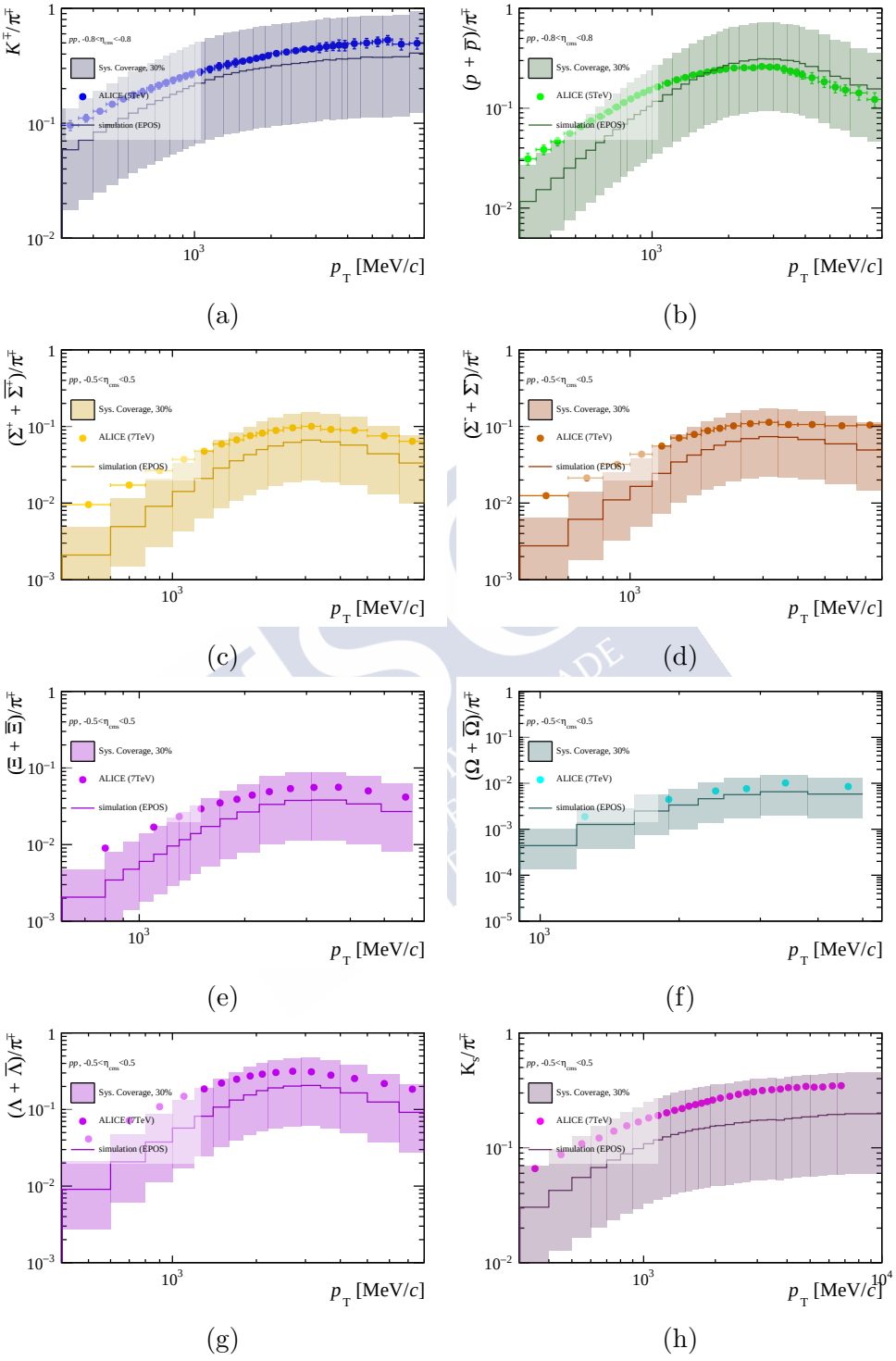
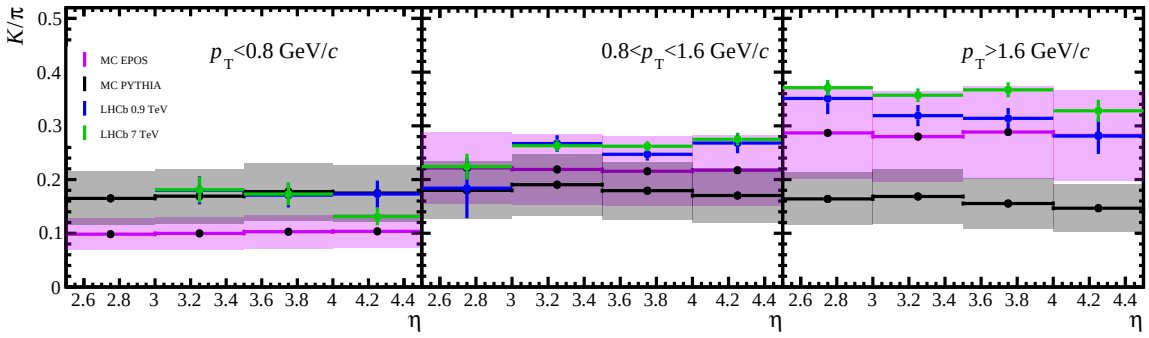
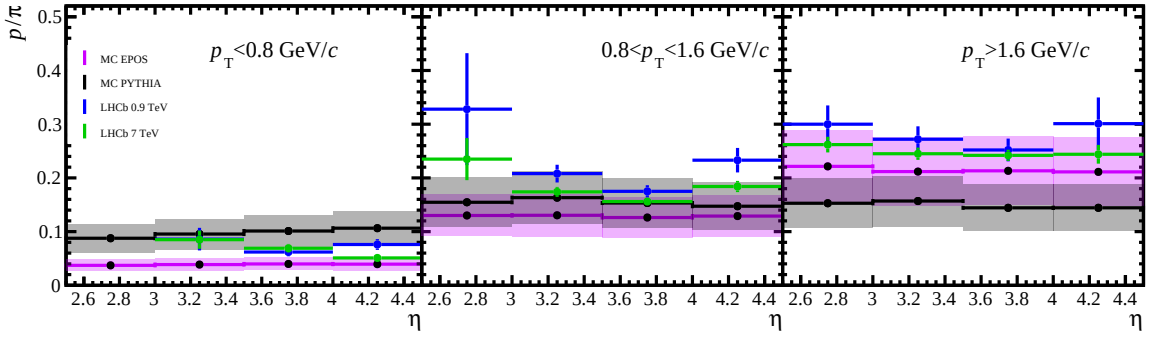


Figure 5.24: Comparison of the hadron/π ratios for pp EPOS simulation and ALICE data [93, 163]. The band corresponds to a 30% variation of the ratio from simulation.



(a)



(b)

Figure 5.25: Comparison of the K/π and p/π ratios for pp collisions at 0.9 TeV and 7 TeV [166] and EPOS and Pythia simulation samples at 5 TeV. A momentum cut of $p > 5\,\text{GeV}/c$ is considered.

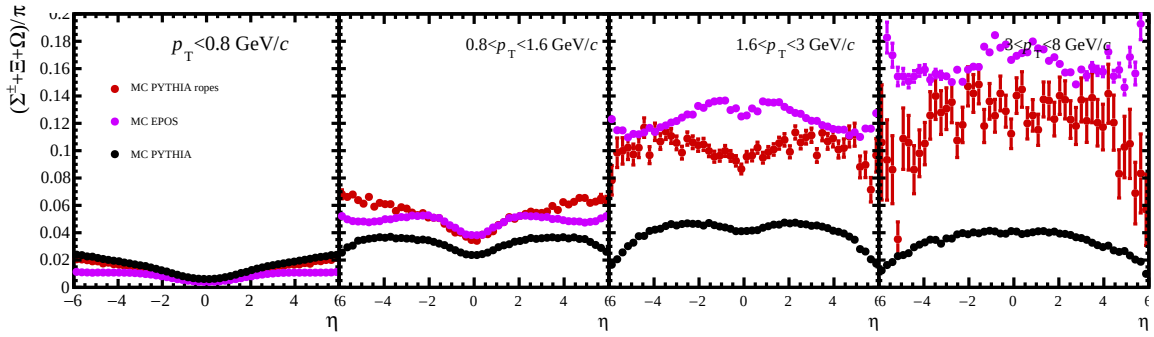


Figure 5.26: $(\Sigma^{\pm} + \Xi^{-} + \Omega^{-})/\pi$ ratios in pp collisions at 5 TeV as predicted by EPOS, Pythia with official LHCb tune and Pythia with string-string interactions.

PYTHIA 8.235 generator (available in Gauss v54r4 version) and with rope hadronisation options [55, 155]. The comparison in the fraction of hyperons/π between the three generators is shown in Fig. 5.26 with respect to η and for four p_{T} bins. It is seen that at high p_{T} the differences between EPOS and PYTHIA can be up to a factor of ten. On the other hand, PYTHIA with rope hadronisation agrees better with EPOS, and any possible difference can be covered by the 30% systematic uncertainty. As conclusion, the relative fractions from the EPOS sample will be used to compute the reconstruction efficiency in *pp* (see Sec. 6.2.2 for more details).

Proton-lead validation. As for *pp* events, the data from the ALICE collaboration in the central region is employed to validate the simulation. In addition to the production of light hadrons, the ALICE collaboration has also measured the production of Ξ^- and Ω^- hyperons in *p*Pb collisions at 5 TeV in the $-0.5 < \eta < 0$ rapidity range [168]. Additionally, Λ and K^0_{S} were measured at 5.02 TeV in $0 < y_{cms} < 0.5$ rapidity range [169]. The $\Sigma^{\pm}$ yields in data can be derived from the Λ yield using the $\Sigma^{\pm}/\Lambda$ ratio from simulation, as was done for *pp* collisions.

The ratio of invariant yields of each hadron with respect to π can be compared with the same ratio in simulation, as shown in Fig. 5.27. In this case, only the composition of the official EPOS sample is explored. Ratios in data are in agreement with the simulation within the $\pm 30\%$ systematic uncertainty for all hadrons.

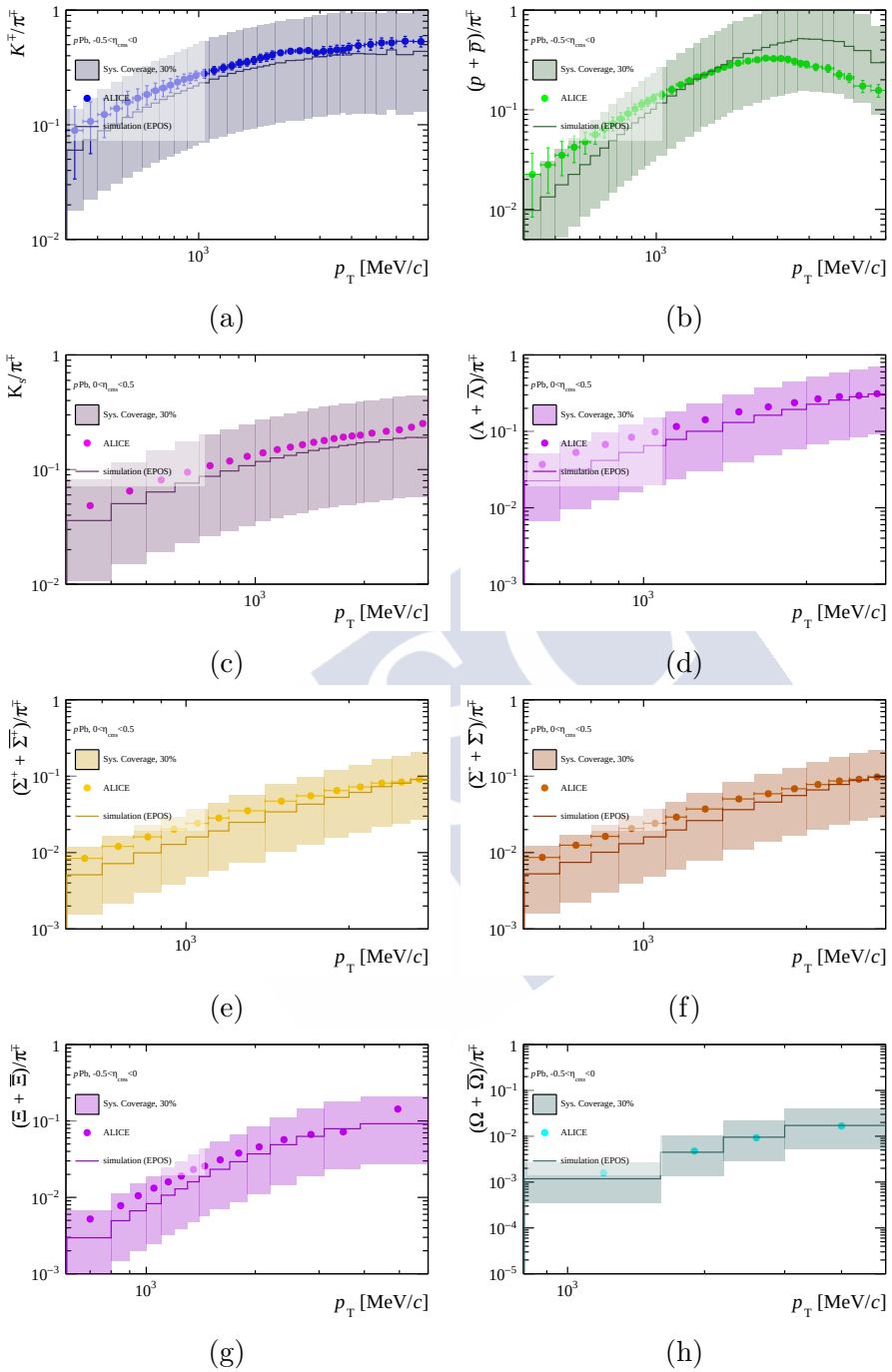


Figure 5.27: Comparison of the hadron/π ratio for Pbp ($-0.5 < \eta_{\rm cms} < 0.5$) and pPb ($0 < \eta_{\rm cms} < 0.5$) collisions in simulation (EPOS) and ALICE data [93, 168, 169]. The band corresponds to a 30% variation of the ratio. Error bars in ALICE data represent the systematic and statistical uncertainty added in quadrature, but they are only displayed in the K/π and p/π ratios.

6

Measurement of the prompt charged particle spectra

The oal of this measurement is to determine the prompt charged particle yield, $N^{\mathrm{ch}}(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$, which is needed to compute the production cross-sections with Eq. 2.12. The starting point is to consider the number of selected candidates, $N_{\mathrm{candidates}}$, obtained after the selection discussed in Secs. 5.1 and 5.3. The prompt charged particle yield can be obtained from the number of candidates as

$$N^{\mathrm{ch}} = N_{\mathrm{candidates}} \frac{P}{\varepsilon_{\mathrm{reco}}\varepsilon_{\mathrm{sel}}(1/\varepsilon_{\mathrm{TM}})} C_{\mathrm{boost}} = \frac{N_{\mathrm{candidates}}}{\varepsilon_{\mathrm{total}}}, \tag{6.1}$$

where P is the purity of the candidate sample, $\varepsilon_{\mathrm{reco}}$ is the reconstruction efficiency, $\varepsilon_{\mathrm{sel}}$ is the selection efficiency, $\varepsilon_{\mathrm{TM}}$ is the truth-matching efficiency and C_{boost} is a correction to the yield due to the boost of the centre-of-mass system in pPb and Pbp collisions. All the corrections can be grouped in the total efficiency $\varepsilon_{\mathrm{total}}$. The purity and every efficiency are $\eta-$ and $p_{\mathrm{T}}-$dependent and must be computed in different (η, p_{T}) bins. The binning choice is detailed in Sec. 6.1.

A first estimation of some of these factors can be obtained using simulation. This will be indicated with the superscript sim. In particular, the reconstruction efficiency, $\varepsilon^{\mathrm{sim}}_{\mathrm{reco}}$, can be computed as

$$\varepsilon^{\mathrm{sim}}_{\mathrm{reco}} = \frac{\#\text{matched long tracks in LHCb}}{\#\text{generated prompt charged particles}}. \tag{6.2}$$

The selection efficiency is defined as

$$\varepsilon_{\text{sel}}^{\text{sim}} = \frac{\#\text{matched and selected long tracks in LHCb}}{\#\text{matched long tracks in LHCb}}. \tag{6.3}$$

Finally, the purity, P^{sim}, can be calculated as

$$P^{sim} = \frac{\#\text{signal candidates}}{\#\text{candidates}}. \tag{6.4}$$

The truth-matching efficiency is a genuine simulation feature that originates from the criteria to pair a particle and a track in simulated data samples. It affects the sample used to compute the selection efficiency. It is defined as

$$\varepsilon_{\text{TM}} = \frac{\#\text{matched and selected long tracks in LHCb}}{\#\text{signal candidates}}. \tag{6.5}$$

For the factors in Eqs. 6.2, 6.3 and 6.4, and throughout Secs. 6.2, 6.3 and 6.4, the discrepancies between data and simulation are studied and final determinations of $\varepsilon_{\text{reco}}$, ε_{sel} and P are measured. The truth-matching efficiency is directly estimated in the simulation sample in Sec. 6.5. The yield correction C_{boost} due to the boost of the centre-of-mass system in pPb and Pbp collisions is addressed in Sec. 6.6 and bin migration effects are studied in Sec. 6.7.

6.1 Binning choice

There are some aspects that should be considered in the binning selection for the measurement of the candidate yield and the correction factors. Ideally, the finest binning could reveal features at a smaller scale in the nuclear modification factors. However, limitations arise from the detector resolution and, mainly, from the size of the data and simulation samples, that have an impact in the relative uncertainties of every bin measurement.

The analysis results are reported in terms of the centre-of-mass pseudorapidity η_{cms}. The relation that links η_{cms} with η is

$$\eta_{\text{cms}} = \eta + y_{boost}.$$

The rapidity shift of the boost to the centre-of-mass in pPb (Pbp) samples is -0.465 (0.465), close to 0.5. This is why an η binning with equally sized bins of 0.5 units is proposed, enabling a good mapping between the centre-of-mass and laboratory system bins. Exceptionally, the last pseudorapidity bin spans from ≈ 4.5 to ≈ 4.8 in the laboratory system. In Tab. 6.1, a summary of the proposed binning scheme in η_{cms} is shown, as well as its correspondence with η in the laboratory system for each configuration.

The p_{T} binning considers that the charged particle multiplicity decreases exponentially with this variable. A set of 14 bins with equal logarithmic range is proposed. The last bin is slightly enlarged to increase its contained sample size.

$\lvert\eta_{cms}\rvert$	$\eta\,(pp)$	$\eta\,(p\mathrm{Pb})$	$\lvert\eta\,(\mathrm{Pb}p)\rvert$
$[1.6, 2.0]$	–	$[2.065, 2.465]$	–
$[2.0, 2.5]$	$[2.0, 2.5]$	$[2.465, 2.965]$	–
$[2.5, 3.0]$	$[2.5, 3.0]$	$[2.965, 3.465]$	$[2.035, 2.535]$
$[3.0, 3.5]$	$[3.0, 3.5]$	$[3.465, 3.965]$	$[2.535, 3.035]$
$[3.5, 4.0]$	$[3.5, 4.0]$	$[3.965, 4.465]$	$[3.035, 3.535]$
$[4.0, 4.3]$	$[4.0, 4.3]$	$[4.465, 4.765]$	$[3.535, 3.835]$
$[4.0, 4.5]$	$[4.0, 4.5]$	–	$[3.535, 4.035]$
$[4.5, 4.8]$	$[4.5, 4.8]$	–	$[4.035, 4.335]$
$[4.8, 5.2]$	–	–	$[4.335, 4.735]$

Table 6.1: Correspondence between the $\lvert\eta_{cms}\rvert$ and η bins for the different considered configurations. The binning schemes used throughout the analysis are shown.

η_{cms} bins in pp	p_{T} bins [GeV/c]
$[2.0, 2.5]$	$[0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$
$[2.5, 3.0]$	$[0.438, 0.570, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$
$[3.0, 3.5]$	$[0.260, 0.338, 0.438, 0.570, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$
$[3.5, 4.0]$	$[0.200, 0.260, 0.338, 0.438, 0.570, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$
$[4.0, 4.3]$	$[0.200, 0.260, 0.338, 0.438, 0.570, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$
$[4.0, 4.5]$	$[0.200, 0.260, 0.338, 0.438, 0.570, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$
$[4.5, 4.8]$	$[0.200, 0.260, 0.338, 0.438, 0.570, 0.740, 0.961, 1.249, 1.622, 2.107, 2.737, 3.556, 4.619, 6, 8]$

Table 6.2: Proposed binning scheme in the (p_{T}, η) plane.

Some regions in the low η and p_{T} range are outside of the fiducial acceptance of the detector. This is induced by the requirement of $p > 2\,\mathrm{GeV}/c$ for a particle to cross the magnet and produce hits in the downstream trackers. The effect is shown in Fig. 6.1. To create these plots a generator level simulation with a single $p > 2\,\mathrm{GeV}/c$ cut was used. The proposed (p_{T}, η) binning scheme, summarised in Tab. 6.2, removes these outside-of-the-acceptance regions.

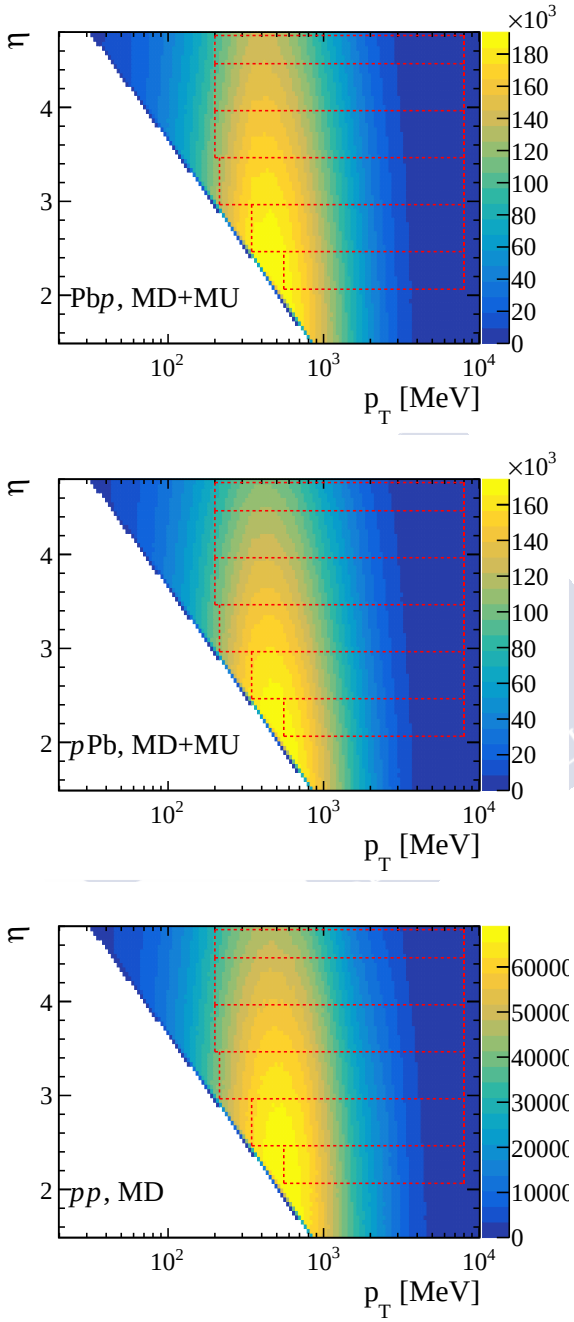


Figure 6.1: Fiducial acceptance of the LHCb spectrometer. In red: pseudorapidity bins and p_{T} range for the different beam configurations. Figures are prepared with simulation including all prompt charged particles with $p > 2\,\mathrm{GeV}/c$.

6.2 Reconstruction efficiency

The reconstruction efficiency, defined in Eq. 6.2, accounts for the prompt charged particles not reconstructed by the spectrometer. It is estimated with simulation by comparing the number of reconstructed tracks matched to a generated prompt charged particle with the total number of generated prompt charged particles. The total number of generated prompt charged particles in the simulation corresponds to all generated particles with `IsPrompt == True`. The valid particles are found using the algorithm described in Sec. 5.3.2. Both the reconstructed track and the generated particle are required to be within the LHCb spectrometer acceptance, see Sec. 5.3.1. Note that the selection requirements detailed in Sec. 5.3.3 do not apply in the equation numerator, since the corresponding correction is considered in the selection efficiency, detailed in Sec. 6.3.

Several corrections to Eq. 6.2 need to be considered to account for differences between data and simulation. For instance, the track reconstruction efficiency is known not to be perfectly reproduced in simulation [132]. A correction factor, $C_{tracking} = \varepsilon^{\text{data}}_{tracking}/\varepsilon^{\text{MC}}_{tracking}$, obtained using a tag-and-probe approach with a calibration sample, is considered. This factor is provided by the LHCb tracking group. Due to different particularities of the proton-lead and proton-proton analyses, they will be discussed separately.

6.2.1 Reconstruction efficiency in proton-lead collisions

Since there is not a specific tracking calibration for the 2013 proton-lead dataset, the 2012 pp data calibration is used instead. This particular sample is chosen because the datatype is 2012 for both cases, and the detector conditions are the same. In Fig. 6.2 the $C_{tracking}$ correction factors used for the Pbp and pPb configurations are shown. The calibration is implemented in (p, η) bins, which need to be translated to the (p_{T}, η) bins discussed in Sec. 6.1. A caveat on this procedure is that the available calibration sample spans over $5 < p < 200\,\text{GeV}/c$. No tracking correction is applied to tracks outside this range.

Since the calibration is prepared with pp data, the dependency of the efficiency with the detector occupancy needs to be considered, since the data-simulation discrepancy in the detector occupancy is larger in pPb and Pbp than it is in pp collisions. By considering the occupancy weights, discussed in Sec. 5.2.1, the occupancy dependence is corrected in simulation and the correction table becomes applicable in the pPb and Pbp samples.

Putting together the kinematic and the occupancy-dependent corrections, the reconstruction efficiency is determined as,

$$\varepsilon_{\text{reco}} = \frac{\sum_{reco\,tracks} C_{tracking,i}(p,\eta) w_i}{\sum_{gen\,parts} w_i}. \tag{6.6}$$

Here, $C_{tracking,i}(p,\eta)$ is the momentum and pseudorapidity dependent tracking correction for the corresponding i-th long track matched with a generated particle and w_i is the occupancy weight assigned to the event of the i-th track or particle.

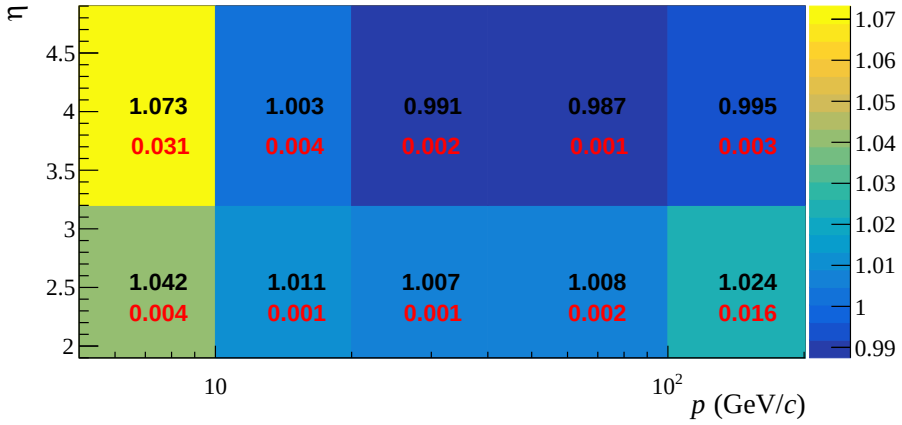


Figure 6.2: Ratio between data and simulation tracking efficiency ($C_{tracking}$) in 2012 pp calibration data. The statistical uncertainty is indicated in red. Data taken from Ref. [170].

To study the dependence of the reconstruction efficiency with the detector occupancy, the reconstruction efficiency is plotted as a function of $N^{\text{clusters}}_{\text{VELO}}$ and $N^{\text{Hits}}_{\text{SPD}}$ in Fig. 6.3. The reconstruction efficiencies of the unweighted and weighted simulation are compared. The value of $\varepsilon_{\text{reco}}$ for different weight sets at a given value of $N^{\text{clusters}}_{\text{VELO}}$ and $N^{\text{Hits}}_{\text{SPD}}$ should be the same. Although this is the case for pPb (Figs. 6.3c and 6.3d) and for $N^{\text{Hits}}_{\text{SPD}}$ in Pbp (Fig. 6.3b), this does not happen for $N^{\text{clusters}}_{\text{VELO}}$ in Pbp. This is because the VELO has positive and negative η acceptance for most events, hence both the forward and backward tracks contribute to $N^{\text{clusters}}_{\text{VELO}}$. This makes the $N^{\text{clusters}}_{\text{VELO}}$ distribution similar between pPb and Pbp, but at the same time, the forward detector occupancy (as $N^{\text{Hits}}_{\text{SPD}}$ for example) very different due to asymmetry in pPb and Pbp. The differences between the $\varepsilon_{\text{reco}}$ values using different weight sets in Fig. 6.3a come from the different description of $N^{\text{Hits}}_{\text{SPD}}$ given by the different sets, as can be seen in Fig. 5.5e.

The corrected efficiencies can be computed by applying the event weight to each reconstructed and generated track of Eq. 6.2. Clearly, the reconstruction efficiency decreases as a function of the occupancy, more notably in Pbp than in pPb events. A similar behaviour is observed for every weights set, with small variations which will be considered when estimating the systematic uncertainty.

The results for the reconstruction efficiency for every (η, p_{T}) bin in the Pbp and pPb configurations are shown in Figs. 6.4a and 6.4b. The weight "set 1" has been used in the computation for both configurations. The results of $\varepsilon_{\text{reco}}$ using the other sets of weights are shown in App. B and in Fig. B.6 for pPb and Fig. B.7 for Pbp. In these figures, error bars show the sum in quadrature of the statistical and the systematic uncertainties. The latter are discussed in Sec. 7.1.

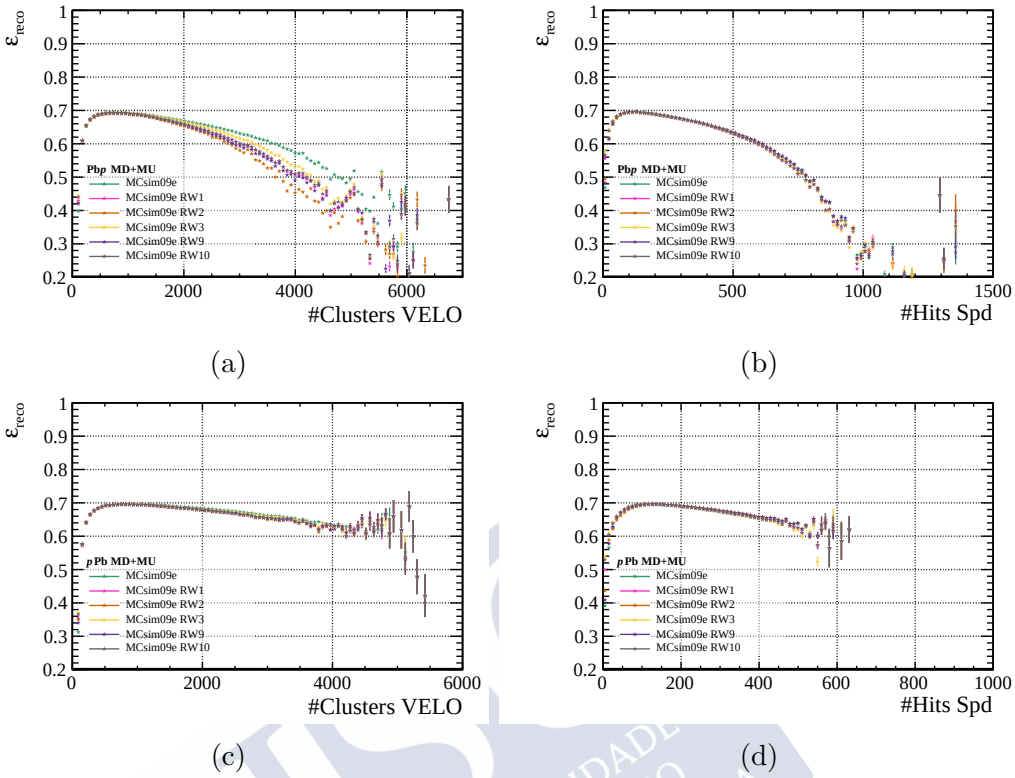


Figure 6.3: Reconstruction efficiency for the Pbp (top) and pPb (bottom) configurations as a function of $N_{\text{VELO}}^{\text{clusters}}$ (left) and $N_{\text{SPD}}^{\text{Hits}}$ (right). Each colour represents the efficiency computed with a different weights set. Error bars include the statistical uncertainties.

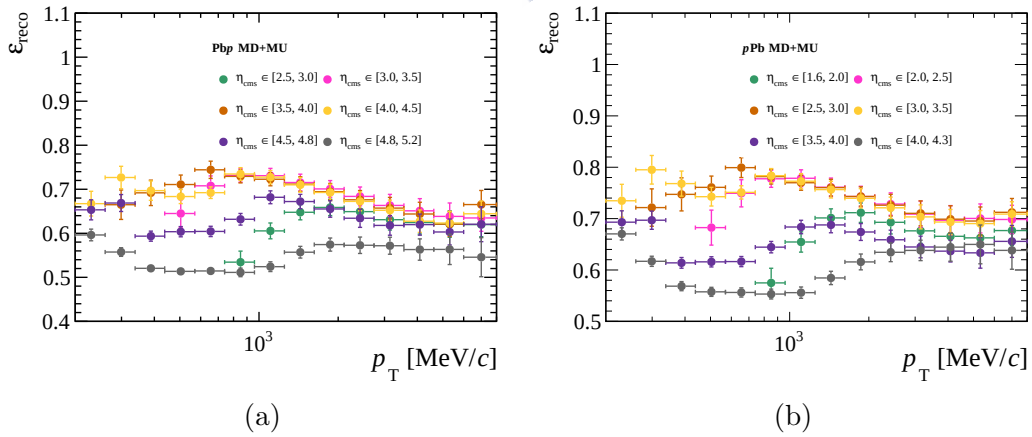


Figure 6.4: Reconstruction efficiency for the Pbp (left) and pPb (right) configurations for the different (η, p_{T}) bins, computed with Eq. 6.6. Datasets with MD and MU are combined. Error bars include statistical and systematic uncertainty.

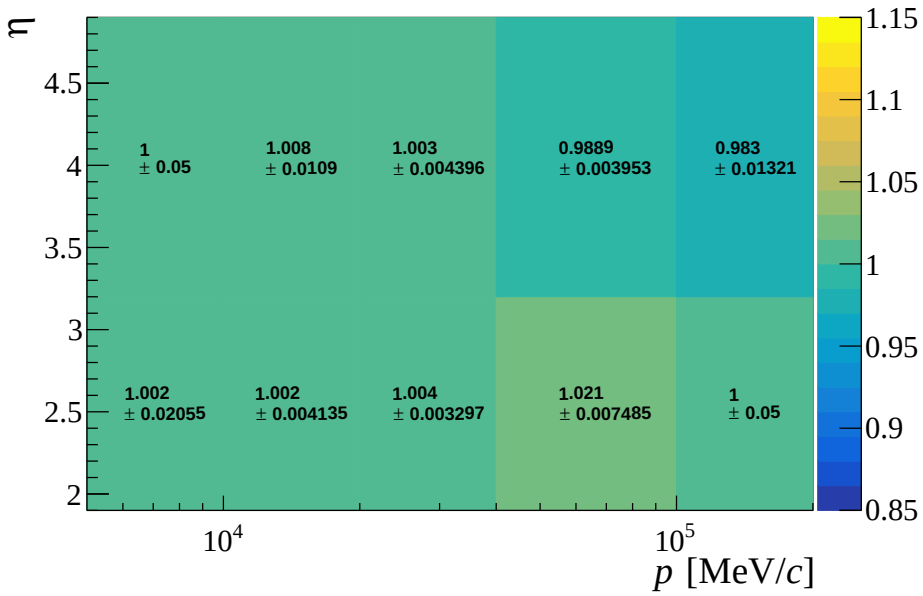


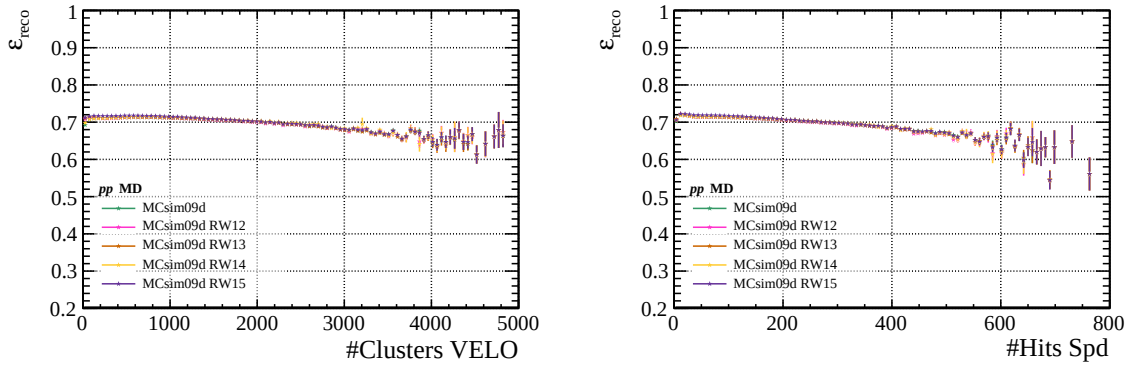
Figure 6.5: Ratio between the tracking efficiency of data and simulation ($C_{tracking}$) for 5 TeV pp 2015. Data from Ref. [170].

6.2.2 Reconstruction efficiency in proton-proton collisions

In pp events, the tracking calibration table from 2015 data at $\sqrt{s} = 5$ TeV is available. The calibration is performed by comparing $B^+ \to J/\psi K^+$ decays between data and the `Sim9b` simulation version. The simulation version of the pp events used in the analysis is `Sim9d`, as shown in Tab. 4.6. However, no differences between these two simulation versions were found. The values of $C_{tracking}(p, \eta)$ from the calibration are shown in Fig. 6.5. For this calibration, the requirement `GhostP` < 0.30 was applied. Therefore, this condition must be also required in the numerator of Eq. 6.2 for the calibration correction factor to be applicable.

The dependency of the reconstruction efficiency as a function of the number of VELO clusters in pp data is shown in Fig. 6.6. The efficiency slightly decreases with the detector occupancy as for proton-lead configurations, although the average event multiplicity is much lower for the pp configuration. For this reason, reproducing the detector occupancy dependency in data is less important for the pp configuration. To verify this assumption, the weights are introduced as was done for pPb and Pbp events. This correction and the $C_{tracking,i}$ factor leave equation 6.2 as

$$\varepsilon_{\text{reco}} = \frac{\sum_{reco\,tracks} C_{tracking,i}(p,\eta) w_i}{\sum_{gen\,parts} w_i}, \tag{6.7}$$



(a) (b)

Figure 6.6: Reconstruction efficiency for pp as a function of $N_{\text{VELO}}^{\text{clusters}}$ (left) and $N_{\text{SPD}}^{\text{Hits}}$ (right), computed with Eq. 6.7. Error bars account for statistical uncertainties.

where $C_{tracking,i}(p,\eta)$ is the p- and η- dependent tracking correction for the corresponding i-th long track matched to a particle, and w_i is the weight of the event that corresponds to the i-th track or particle. In Fig. 6.7, the reconstruction efficiency is represented without weighting the simulation and with the different sets of weights. The plot confirms the small effect of the detector occupancy in the efficiency by showing agreement among the results obtained with the different sets of weights.

However, an additional correction to $\varepsilon_{\text{reco}}$ needs to be introduced in the pp sample. As it was seen in Sec. 5.4.3, the pp simulation generated with the LHCb tune of PYTHIA, does not reproduce well the relative prompt charged particle abundances in data. This has an effect in the reconstruction efficiency because each particle species has a different reconstruction efficiency. The reconstruction efficiency per particle can be computed with equation 6.7 by requiring a given truth ID. The $\varepsilon_{\text{reco}}$ per particle is shown in Fig. 6.8. Different behaviours are observed. On the one hand, pions, kaons and protons, which constitute the bulk of the sample, have similar efficiencies around ≈ 0.7. The small differences originate in the different interaction with the detector material and in the fact that pions and kaons can decay before leaving a long track, while protons cannot. On the other hand, hyperons have very small reconstruction efficiency since they mostly decay before reaching the downstream tracking stations. Their efficiency only increases slightly for very high p, as such hyperons have a boost high enough to cross the LHCb magnet. The efficiency for muon and electrons is dominated by the electron efficiency, but in any case they do not play an important role due to their low relative abundance.

Given this situation, the particle species composition in the official simulation presents two issues to estimate $\varepsilon_{\text{reco}}$: the underestimation of the hyperon/π fraction and the underestimation of the K/π fraction at high p_{T}. PYTHIA underestimates these fractions by an order of magnitude as compared to ALICE data, EPOS simulation and PYTHIA

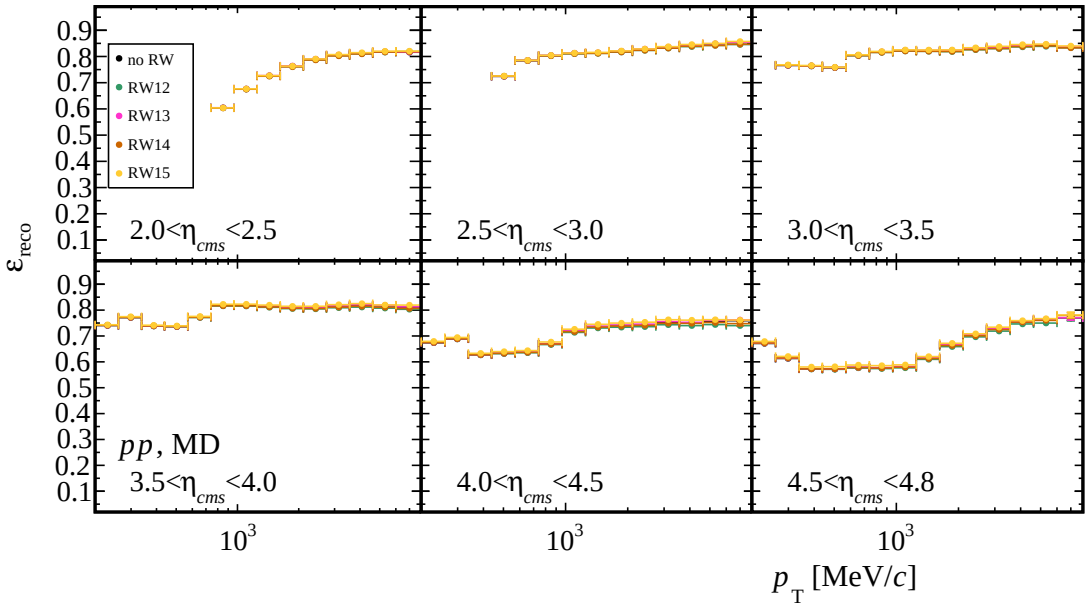


Figure 6.7: Dependency of the reconstruction efficiency as a function of p_{T} in every η_{cms} bin for the different sets of weights in pp data.

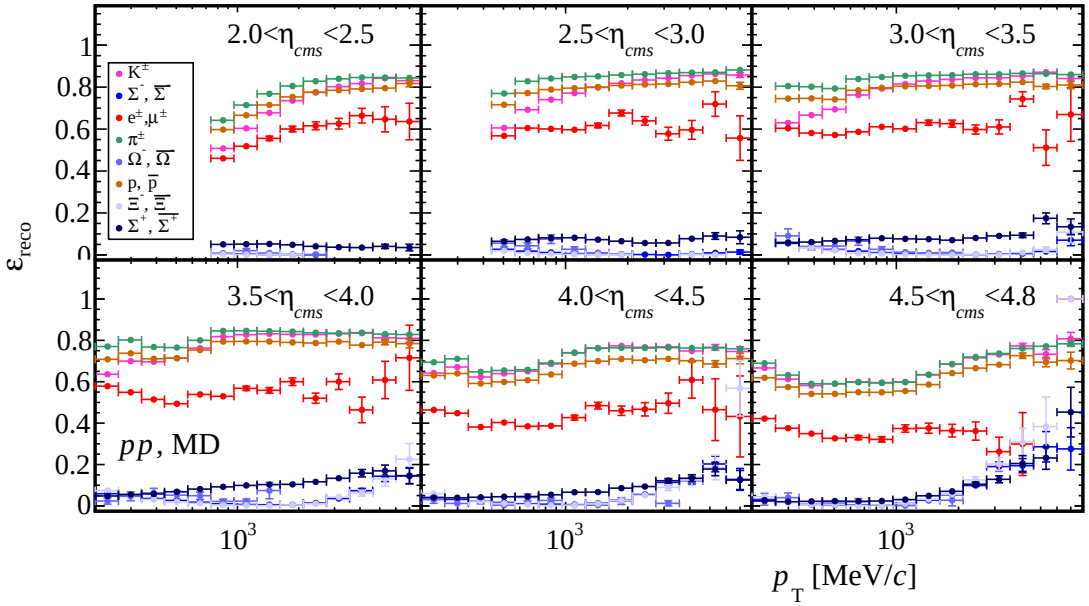


Figure 6.8: Reconstruction efficiency as a function of p_{T} in every η_{cms} bin for all prompt charged particle species.

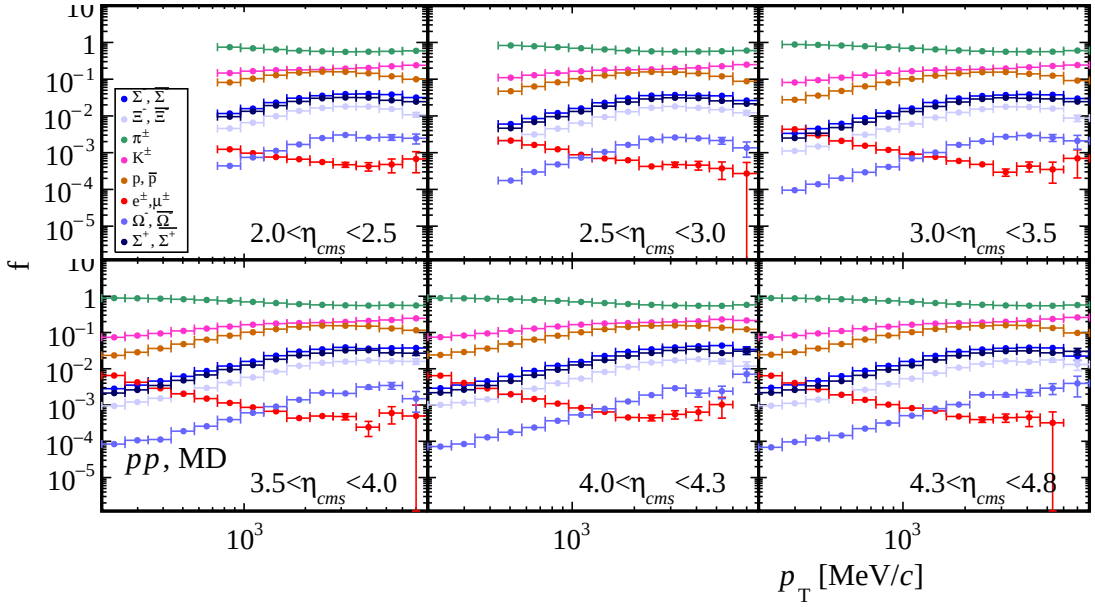


Figure 6.9: Relative abundances of all prompt charged particle species in EPOS pp simulation.

simulation with hadronisation ropes.

To solve these issues, the total reconstruction efficiency can be expressed in terms of the individual $\varepsilon_{reco,p}$ per particle:

$$\varepsilon_{\text{reco}} = \sum_p f_p \varepsilon_{reco,p}, \quad \text{with}\, p = \pi^-, K^-, p, e^-, \mu^-, \Xi^-, \Sigma^+, \Sigma^-, \Omega^-, cc. \tag{6.8}$$

Here, f_p is the abundance of particle p relative to the total number of prompt charged particles, and $\varepsilon_{reco,p}$ is the reconstruction efficiency from Eq. 6.7 for particle p. Since the EPOS generator has been validated with data and gives a reasonable description of the relative abundances, f_p, in Eq. 6.8 are computed with a private EPOS pp sample as those are independent of the reconstruction and acceptance of the LHCb detector. These fractions are shown in Fig. 6.9. The official Pythia sample is used to extract $\varepsilon^p_{\text{reco}}$ (Fig. 6.8).

The effect of considering EPOS instead of Pythia for the particle abundances is shown in Fig. 6.10. The difference between both results is significant, up to 8% for high p_{T} due to the hyperons and kaons discrepancy.

Finally, the overall reconstruction efficiency for the pp sample is shown in Fig. 6.11. In Fig. 6.12, $\varepsilon_{\text{reco}}$ is compared for the three beam configurations.

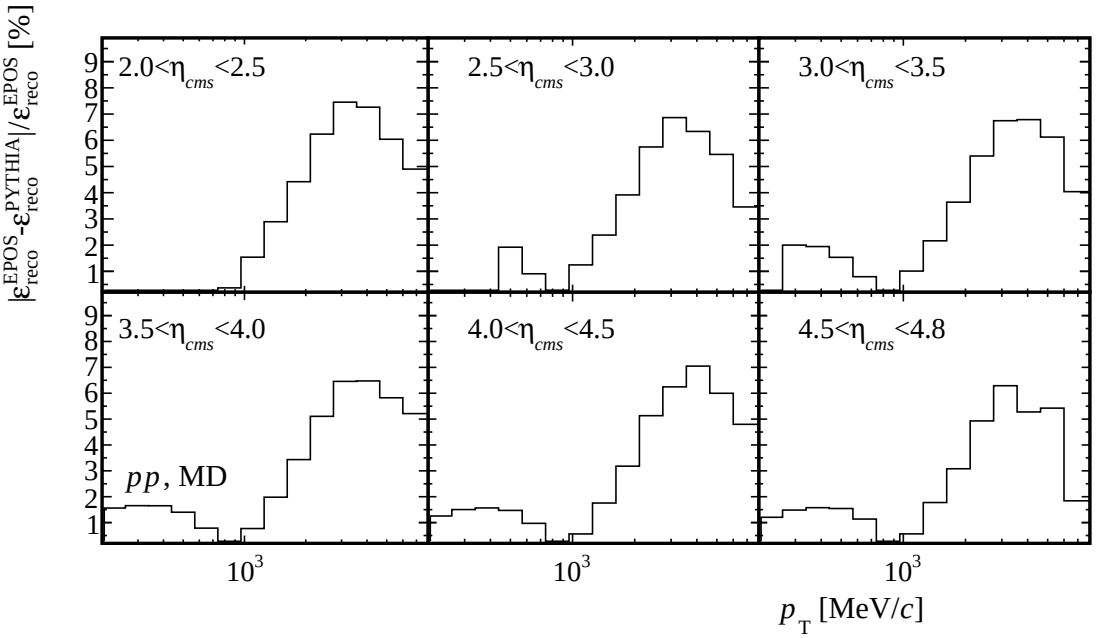


Figure 6.10: Relative difference between $\varepsilon_{\text{reco}}$ computed with Pythia and EPOS particle fractions (f_p).

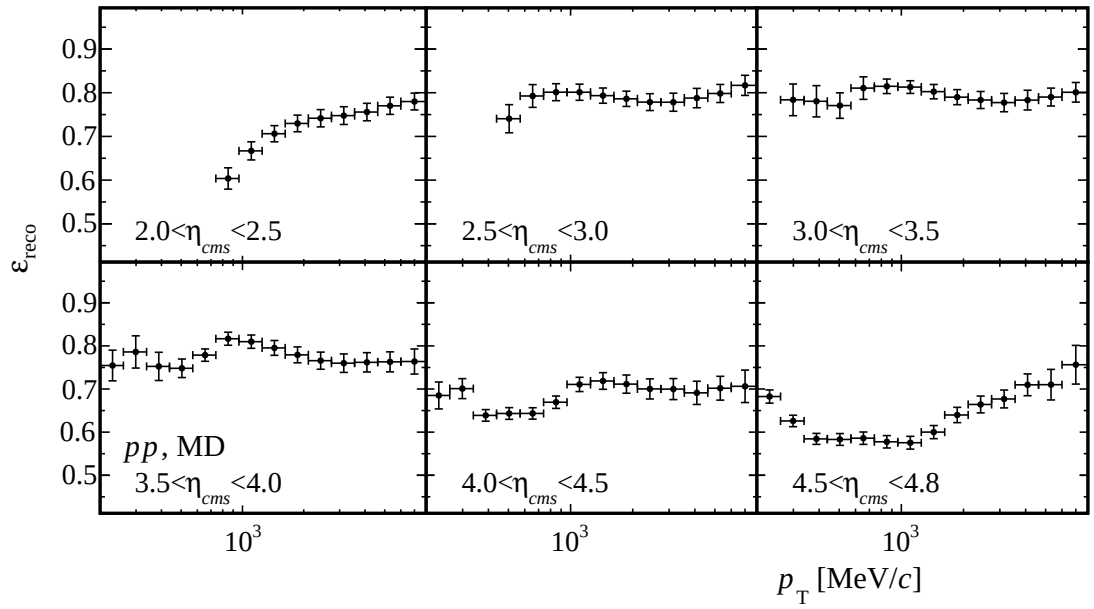


Figure 6.11: Overall reconstruction efficiency for the *pp* sample. The error bars correspond to statistical and systematic uncertainties (see Sec. 7.1 for more details) added in quadrature.

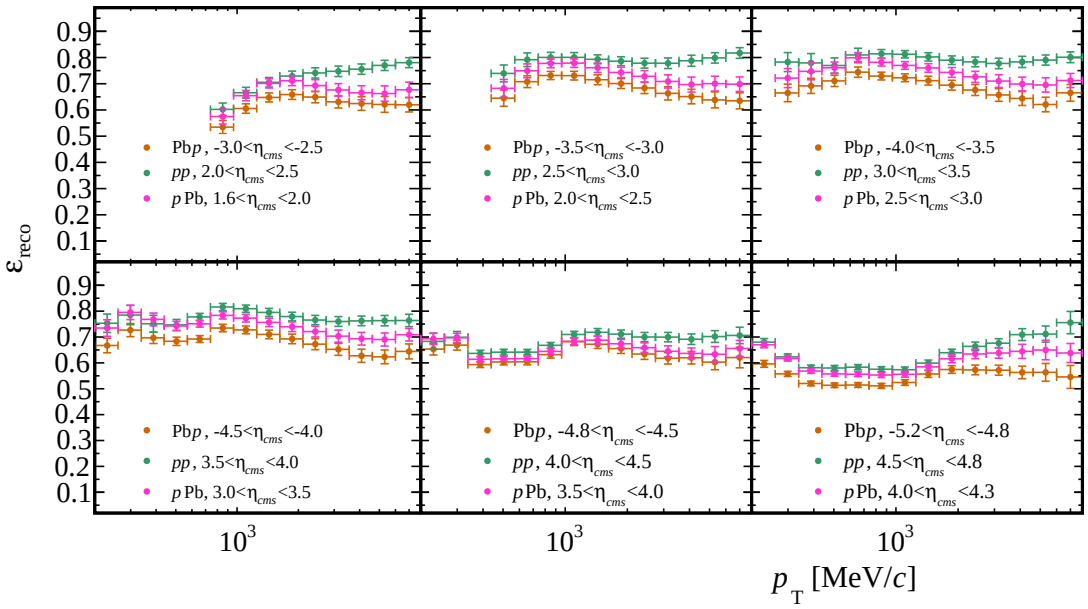


Figure 6.12: Reconstruction efficiency $\varepsilon_{\text{reco}}$ for Pbp, pPb and pp data.

6.3 Selection efficiency

The selection efficiency is obtained from simulation as the number of matched tracks passing the selection criteria divided by the total number of matched long tracks within the LHCb detector acceptance, as described in Eq. 6.3. This efficiency corrects $N_{\text{candidates}}$ accounting for real prompt charged particles reconstructed by the spectrometer but discarded by the selection requirements. The considered selection includes the requirements in `GhostP`, `pseudoIP` and non shared VELO tracks.

As already mentioned, simulation in proton-lead data does not accurately reproduce the detector occupancy. This has an impact in the selection variables, particularly in ghost probability distributions, as seen in Figs. 5.21e and 5.21f for proton-lead and also in Fig. 5.22c for the pp data. These figures show that the weighting is not sufficient to fully correct the data-simulation discrepancies in `GhostP`. Some discrepancy in the `pseudoIP` distribution is also seen for the three datasets in Figs. 5.22d, 5.21h and 5.21g. Additionally, the tracking-efficiency calibration tables disregard the selection efficiency, and the considered `GhostP` requirement is tighter than what is typically considered in other LHCb analysis. All in all, the selection efficiency needs to be studied in detail with a data-driven calibration procedure.

A correction factor $C_{sel} = \varepsilon_{\text{sel}}^{\text{data}}/\varepsilon_{\text{sel}}^{\text{sim}}$ using a calibration sample is proposed, in a similar way to the calibration of the reconstruction efficiency. Once this factor is obtained, the

selection efficiency can be computed as

$$\varepsilon_{\text{sel}} = \frac{\sum_{sel\,tracks} C_{sel,i}}{N_{reco\,tracks}}. \tag{6.9}$$

Note that in this equation only tracks matched with a MC particle are considered. In the following section the details to obtain C_{sel} are discussed. A calibration procedure with $\phi \to K^{\pm}K^{\mp}$ decays is explained in Sec. 6.3.1, to be used in the analysis.

6.3.1 Calibration with $\phi \to K^{\pm}K^{\mp}$ decays

The selection efficiency is measured both in data and simulation from background subtracted samples of prompt charged particle candidates using a tag-and-probe method, and the correction factor C_{sel} is computed. The background subtracted samples are obtained from the decay products of a resonance whose invariant mass is employed to discriminate the signal from the background. In this analysis, the chosen decay is the $\phi(1020) \to K^{\pm}K^{\mp}$ mode. The $\phi(1020)$ meson is abundantly produced in hadronic collisions and its decay into two oppositely charged kaons has a large branching fraction: $\mathcal{BR}(49.2 \pm 0.5)\%$ [171]. Also, since the $\phi(1020)$ meson is a very short lifetime resonance, $\tau \sim 10^{-22}\,\text{s}$, its decay products are considered as prompt charged particles if the resonance was produced in the hadronic interaction, or if its mother particle has a lifetime lower than 30 ps (see definition in Sec. 2.5.1. The $\phi(1020)$ meson is also included in the minimum bias simulation, both in proton-lead and *pp* events.

Positively charged kaons (K^+) of the `StdAllNoPIDsKaons` type are used as tag. The tag purity is increased by requiring `ProbNNK` > 0.5. The probe sample is constituted by every long track in the `TES["Rec/Track/Best"]` container and within the LHCb acceptance as explained in Sec. 5.3.1). The kaon pair is required to have an invariant mass within the $1000 < M(K^{\pm}K^{\mp}) < 1040\,\text{MeV}/c^2$ range and a distance of closest approach (`DOCA`) of less than 0.2 mm. The procedure is repeated by taking K^- candidates as tags and K^+ candidates as probes, in order to increase the sample size and to avoid charge biases. The selection is summarised in Tab. 6.3.

After identifying the $\phi(1020)$ contribution, the probe kaons constitute a sample of unbiased candidates with respect to the selection. Therefore, the selection efficiency can

$K^{\pm}$ (tag)	$K^{\mp}$ (probe)	$\phi(1020)$ (combination cuts)
`StdAllNoPIDsKaons`	`TES["Rec/Track/Best"]`	$M(K^{\pm}K^{\mp}) < 1040\,\text{MeV}/c^2$
`ProbNNK` > 0.5	Long tracks	$M(K^{\pm}K^{\mp}) > 1000\,\text{MeV}/c^2$
	In LHCb acceptance	`DOCA`$(K^{\pm}, K^{\mp}) < 0.2$

Table 6.3: Summary of the selection requirements applied to the $K^{\pm}$ (tag), $K^{\mp}$ (probe) and $\phi(1020)$ candidates.

be extracted from the number of probe kaons that pass the selection as,

$$\varepsilon_{\rm sel} = \frac{N(K^{\mp})|_{selected}}{N(K^{\mp})|_{all}} = \frac{N(\phi(1020))|_{selected}}{N(\phi(1020))|_{all}}, \tag{6.10}$$

where $N(\phi(1020))|_{selected}$ is the signal yield considering the selection requirements to the probe kaon, and $N(\phi(1020))|_{all}$ is the signal yield without additional requirements to the probe kaon.

The $\phi(1020)$ signal yield is extracted by performing a binned fit to the invariant mass of the kaon pair. The signal shape is modelled with a Voigtian function, which is the convolution of a Breit-Wigner distribution for the resonance and a Gaussian distribution that accounts for the detector resolution. The values of the mass and width in the Breit-Wigner distribution are fixed to the values provided by the PDG [171]. The width of the Gaussian is allowed to float in the fit when no selection is applied to the probed kaon, and the result of this fit is used to fix the width when fitting the signal applying the selection. For the pPb and Pbp simulation an abnormally reduced width for the $\phi(1020)$ resonance is observed. By checking the MC truth information it was found that the $\phi(1020)$ resonance width is null in the EPOS generator. For this reason, in proton-lead collisions, the observed width in simulation exclusively originates from the detector resolution. This is not expected to have an impact in the ratio of Eq. 6.10, but requires to set and fix the width value to zero in the fit of these data.

The background is modelled with the function

$$f(M) = q \cdot m_b^{a_1} \cdot e^{-a_2 \cdot m_b}, \tag{6.11}$$

where q is the momentum of a daughter in the resonance rest frame, a_1 and a_2 are free parameters in the fit, and $m_b = M(K^+K^-) - 2m_K$. The yields of signal and background are left free in the fit.

The results of the signal extraction are shown in Fig. 6.13. In each plot the data and the fit results are displayed before (blue line and green dots) and after (red line and black dots) the selection requirements. It is clear that the selection requirements remove an important background fraction from fake tracks, which is seen in the reduction of the background contribution after the fit. The background is specially reduced in the Pbp samples. For the proton-lead datasets, MD and MU configurations are merged to increase the sample size, since the efficiencies obtained with both polarities are compatible. Eq. 6.10 allows to estimate the selection efficiency in data, $\varepsilon_{\rm sel}^{\rm data}$, and in simulation, $\varepsilon_{\rm sel}^{\rm sim}$, to obtain the correction factor for the efficiency $C_{sel} = \varepsilon_{\rm sel}^{\rm data}/\varepsilon_{sel}^{sim}$.

A critical part of the tag-and-probe method is that the selection described in Tab. 6.3 must not bias the sample of probe kaons, which should cover the $(\eta, p_{\rm T})$ ranges of the candidates sample. Due to the different acceptance efficiency of a single track and a track originating in a two-body decay, the $(\eta, p_{\rm T})$ distributions are expected to be different, since for the two body decay both tracks must be within the detector acceptance. To minimise this bias, the correction to the efficiency is computed in $(\eta^{K^{\mp}}, p_{\rm T}^{K^{\mp}})$ bins. By

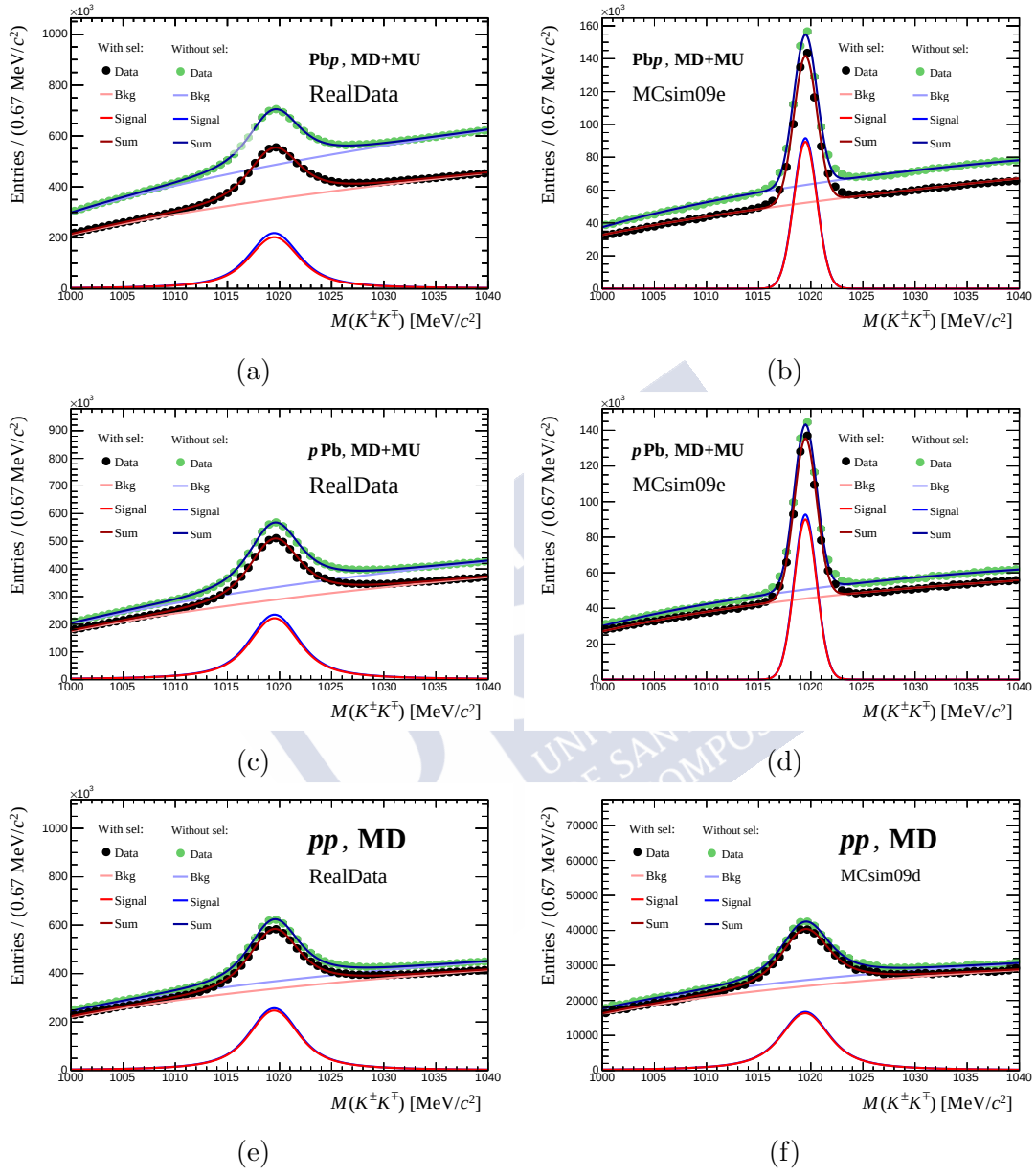


Figure 6.13: Signal extraction of $\phi(1020)$ meson candidates for every configuration. Figures include data before and after the selection requirements, along with the corresponding fits in blue and red, respectively. Statistical uncertainties in data are imperceptible due to the large statistics of the sample.

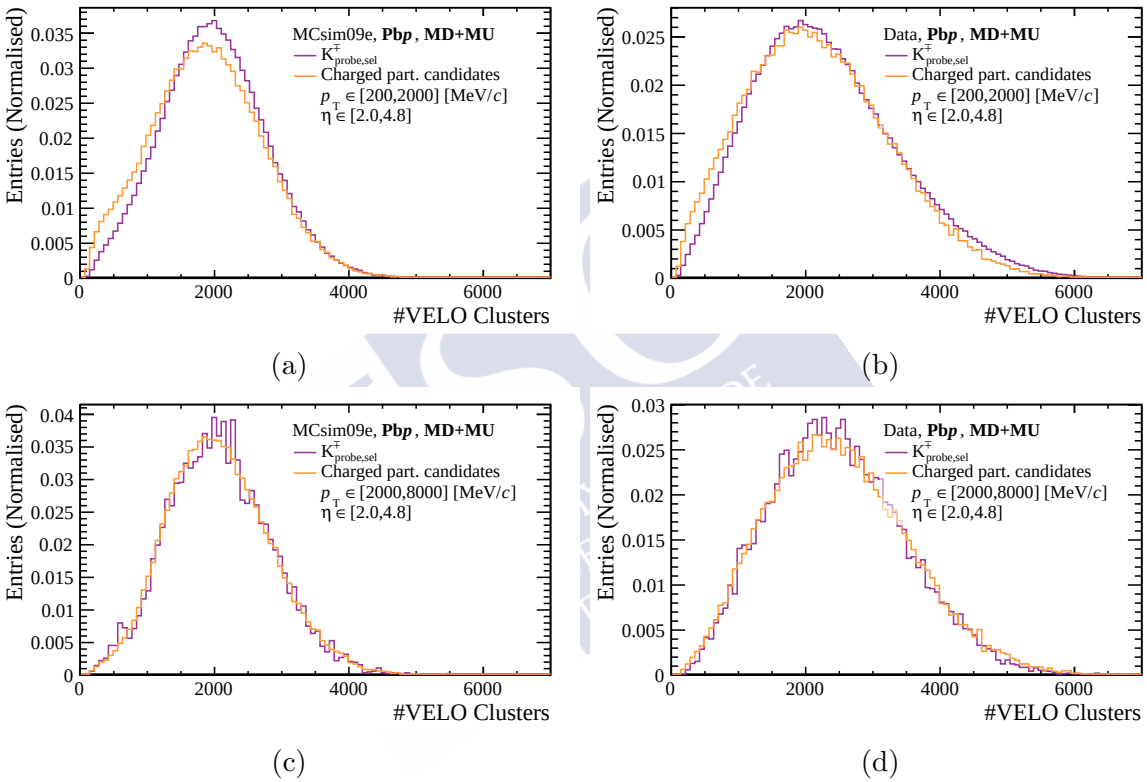


Figure 6.14: Comparison of the number of VELO cluster distributions for charged particle candidates and selected probe kaons in the Pbp configuration for simulation (left) and data (right) for $200 < p_{\mathrm{T}} < 2000\,\mathrm{MeV}/c$ (top) and $2000 < p_{\mathrm{T}} < 8000\,\mathrm{MeV}/c$ (bottom).

binning the correction a better description of the selection efficiency is achieved, since the behaviour of low-p_{T} tracks is different from that of high-p_{T} tracks. The bin sizes are chosen such that they all contain a significant amount of candidates. The two-body invariant mass fits for each bin are summarised in App. B.

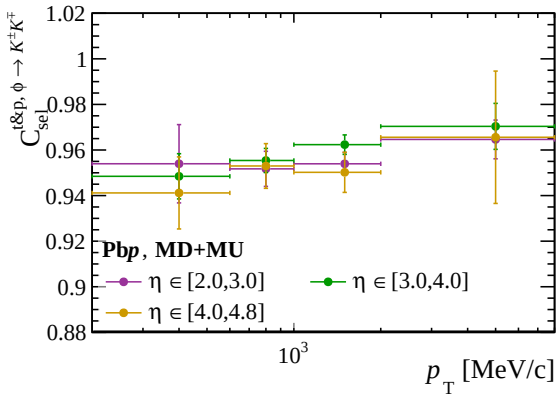
Another relevant issue originates in the event occupancy distributions. The detector occupancy for charged particles candidates and for probe kaons need to be similar so the C_{sel} factor can be extrapolated. The average occupancy could be higher for the probes as two reconstructed tracks are required in the selection. Fig. 6.14 shows the comparison for candidates and probes for low p_{T} (Figs. 6.14a and 6.14b) and for high p_{T} (Figs. 6.14c and 6.14d). The distributions for probe kaons are obtained using the sPlot technique to extract the background in Fig. 6.13 [172]. An excellent agreement between the distributions is seen both in data and simulation and for low and high p_{T}. It is worth noting that a small background contribution is present in the charged particle candidate distributions since the selection requirements do not remove all the background, however the comparison is still fair because after the selection the contribution from background is very small. The same distributions for pPb and pp are shown in the appendix in Fig. B.14 and Fig. B.15, respectively. The agreement between the distributions is very good as well for those datasets.

The distributions of probe kaons with respect to η, p_{T}, `pseudoIP` and `GhostP` are shown in the appendix in Fig. B.16 (Pbp), Fig. B.17 (pPb) and Fig. B.18 (pp). The distributions of probed kaons cover well the region of $2 < \eta < 4.8$ and $200 < p_{\mathrm{T}} < 8000\,\mathrm{MeV}/c$ for all samples. The `pseudoIP` distribution is similar to the distribution from prompt charged particles in Fig. 5.13 (blue markers). The same occurs for `GhostP` distributions compared to the prompt charged particle signal in Fig. 5.9.

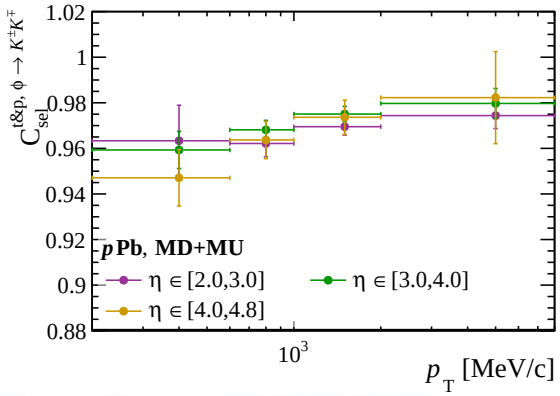
Additionally, a correlation between the variables of the tagged and probed kaons could induce a bias in the probe kaon sample distributions and propagate to the efficiency estimation. To estimate the possible bias, the selection imposed to the tag kaon from Tab. 6.3 is modified by changing the `ProbNNK` and `DOCA`$(K^{\pm}, K^{\mp})$ requirements. The values 0.6, 0.7, 0.8 and 0.9 for `ProbNNK` and 0.17 mm, 0.14 mm and 0.10 mm for `DOCA`$(K^{\pm}, K^{\mp})$ are tested by repeating the fits showed in Fig. 6.13. No kinematic binning is considered for this test to reduce the statistical uncertainty. The effect in the ratio of efficiencies in data and simulation is negligible compared to the systematic uncertainties quoted in Sec. 7.2.

The results for C_{sel} are presented in Fig. 6.15 including the statistical uncertainty from the signal yields of the fit and the systematic uncertainty discussed in Sec. 7.2. The deviation from unity of the correction factor is larger for Pbp data (Fig. 6.15a), followed by pPb data (Fig. 6.15b). The correction factor is compatible with unity for pp data (Fig. 6.15c).

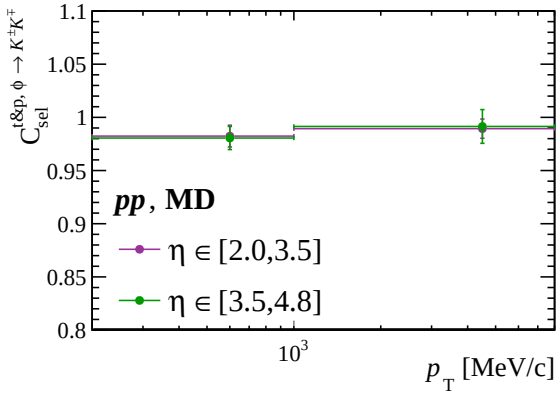
These results reflect that the selection efficiency should decrease with the event multiplicity, and it is underestimated in Pbp and, to a lesser extent, in pPb simulation events. Thus, the selection efficiency is lower in data than in simulation. The correction factor approaches unity with growing p_{T} for every configuration, which reflects than the selection efficiency for high-p_{T} tracks is less multiplicity-dependent. No significant η dependence



(a)



(b)



(c)

Figure 6.15: Correction factor of the selection efficiency for the Pbp, pPb and pp configurations computed with the tag-and-probe method using kaons from $\phi(1020) \to K^{\pm}K^{\mp}$ decays.

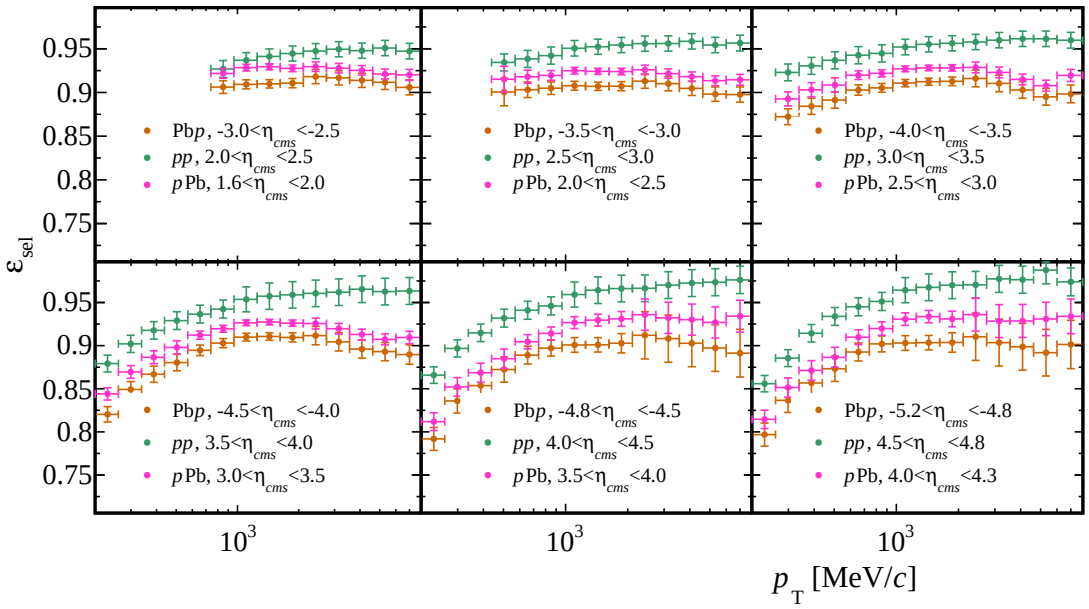


Figure 6.16: Selection efficiency for the Pbp, pPb and pp configurations computed with Eq. 6.9. The error bars include the statistical and systematic uncertainty.

is observed within the uncertainties. The number of (η, p_{T}) bins is less than for the final result (Sec. 6.1), in order to reduce statistical uncertainties due to the available dataset. However, a very smooth variation of the correction value is obtained for the three configurations. Only a small global shift in pPb and Pbp is seen. Finer binning configurations of C_{sel} have been tested and no significant discrepancies with the values on Fig. 6.15 have been observed.

The value of C_{sel} is used to correct the selection efficiency computed with the MC truth information as shown in Eq. 6.9. Finally, the selection efficiency obtained with these correction factors is shown in Fig. 6.16.

6.4 Background subtraction

After the selection requirements in `GhostP`, `pseudoIP` and the shared VELO track removal, the distribution of candidates still contains a contribution from background tracks. In this section, the procedure to estimate this contribution is addressed.

The purity in simulation, P^{sim}, can be obtained with Eq. 6.4 by identifying the background candidates in the sample. As mentioned in Sec. 5.3.2, the main background sources are fake tracks, secondary particles, and, to a lesser extent, clone tracks. Being $f^{\mathrm{sim}}_{\mathrm{fake}}$, $f^{\mathrm{sim}}_{\mathrm{sec}}$ and $f^{\mathrm{sim}}_{\mathrm{clone}}$ the fractions of each of these background components in simulation, the signal

purity can be obtained as,

$$P^{\text{sim}} = 1 - (f^{\text{sim}}_{\text{fake}} + f^{\text{sim}}_{\text{sec}} + f^{\text{sim}}_{\text{clone}}). \tag{6.12}$$

These background fractions can be directly computed since the nature of the candidates in simulation is known. A signal candidate consists in a selected long track matched with a MC particle that meets the prompt requirements, as defined in Sec. 2.5. The background categories were already defined in Sec. 5.3.2.

In simulation, the fraction of fake tracks can be computed as,

$$f^{\text{sim}}_{\text{fake}} = \frac{\sum_i N_{\text{non-matched cand},i}\, w_i \varepsilon_{\text{TM}}}{\sum_i N_{\text{cand},i} w_i}, \tag{6.13}$$

where ε_{TM} is the truth-matching efficiency and w_i is the weight of the i-th event that corrects the detector occupancy dependence, discussed in Sec. 5.2. The ε_{TM} factor is introduced because the truth matching algorithm is not fully efficient. When the truth-matching algorithm fails, a fake track is accounted for the candidate that was not matched to its corresponding MC particle. The procedure to extract this efficiency is discussed in detail in Sec. 6.5.

The fraction of secondary particles can be written as,

$$f^{sim}_{sec} = \frac{\sum_i N_{non-prompt\, cand} w_i}{\sum_i N_{cand} w_i}. \tag{6.14}$$

Finally, the fraction of duplicated candidates in simulation can be computed as,

$$f^{\text{sim}}_{\text{clone}} = \frac{\sum_i N_{clone\, cand} w_i}{\sum_i N_{cand} w_i}. \tag{6.15}$$

The dependency of the background fractions with the detector occupancy can be studied. In Fig. 6.17, each of the background components is represented with respect to the number of VELO clusters and the number of SPD hits of the event for the three different beam configurations. The fraction of fake tracks in Figs. 6.17a and 6.17b rises strongly with the event occupancy for the three beam configurations, specially for Pbp. For Pbp and pPb, a similar tendency in the number of SPD hits is seen, apart from the different cutoff which is due to the different reach in event multiplicity of the two samples. This is not the case for $N^{\text{clusters}}_{\text{VELO}}$, where Pb$p$ has a steeper increase with respect to pPb. This was seen already in Fig. 5.8c before considering any selection requirement to the reconstructed tracks. The explanation of this different behaviour is the VELO acceptance, which covers part of the backward region, giving an additional contribution to the multiplicity in pPb that comes from the backward part of the detector, which sees a Pbp collision. Another possible explanation could be a partial saturation-like effect of the VELO detector for the events with the highest occupancies in Pbp, which should not observed in the SPD. This can be checked in Fig. 6.18, where the correlation between

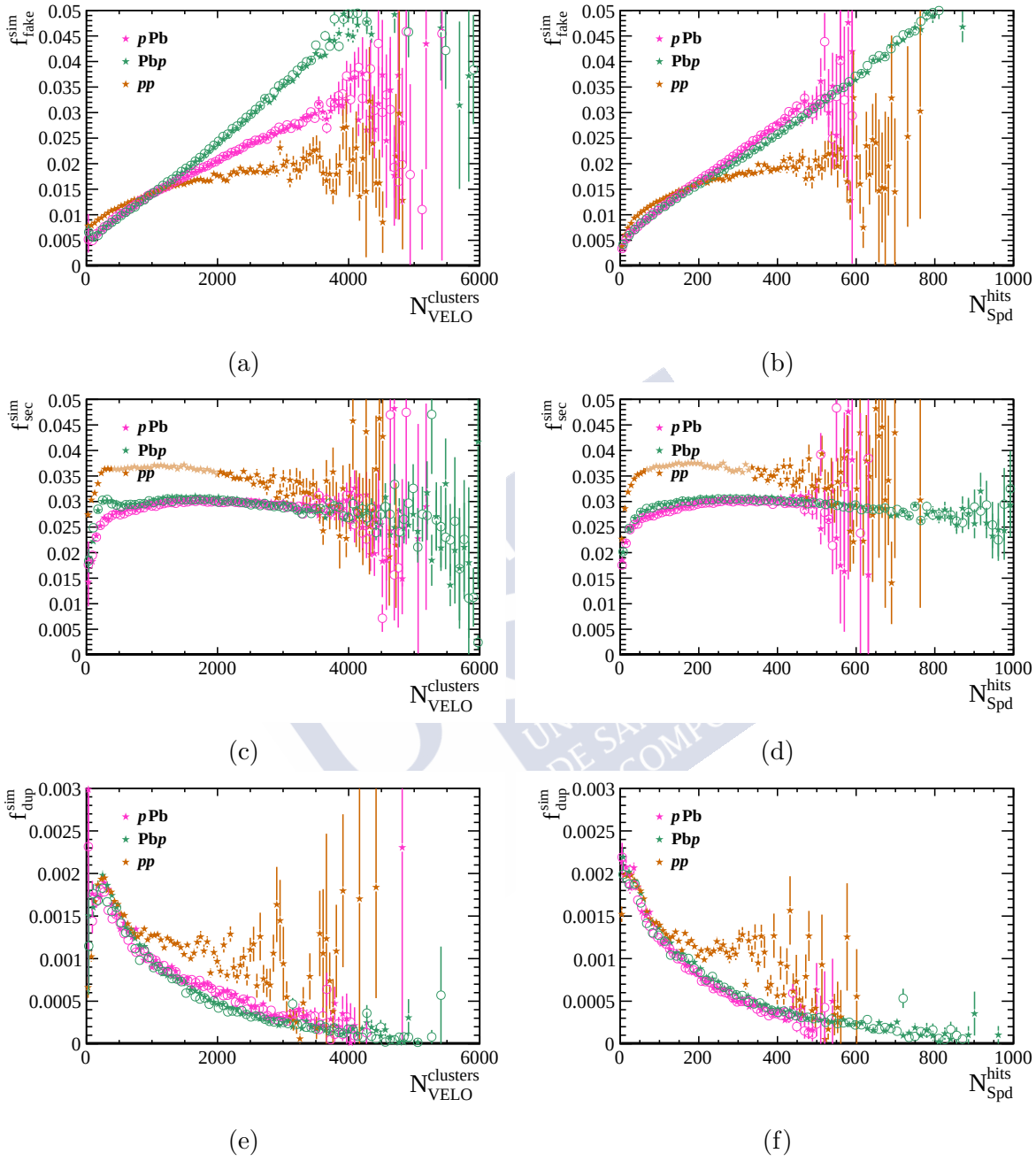
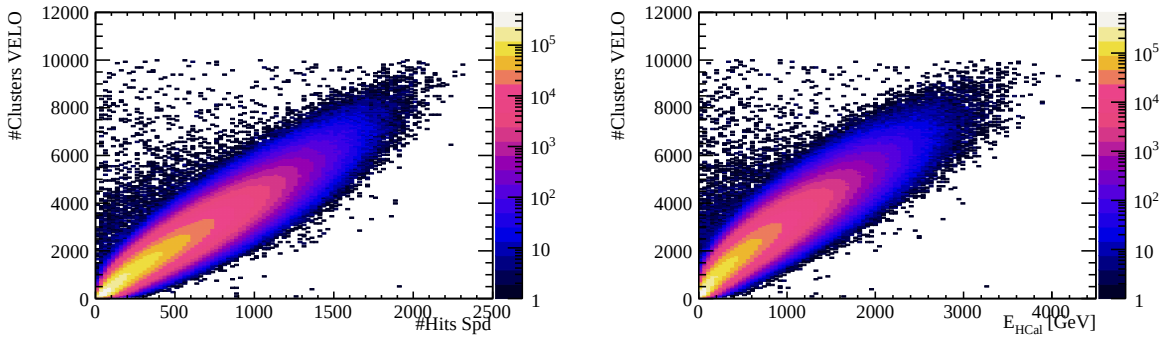


Figure 6.17: Dependency of the fraction of fakes (top), secondaries (middle) and clones (bottom) in the candidate sample of simulation with respect to the number of VELO clusters (left) and the number of SPD hits (right). Circles represent the MD sample while stars represent the MU sample. Weight sets 1 and 12 are considered for proton-lead and *pp*, respectively.



(a) (b)

Figure 6.18: Correlation between $N_{\mathrm{VELO}}^{\mathrm{clusters}}$ and $N_{\mathrm{SPD}}^{\mathrm{Hits}}$ (left) and between $N_{\mathrm{VELO}}^{\mathrm{clusters}}$ and the total energy collected at the HCAL. Data corresponds to Pbp MD configuration.

$N_{\mathrm{VELO}}^{\mathrm{clusters}}$ and $N_{\mathrm{SPD}}^{\mathrm{Hits}}$ and the energy collected in the HCAL is presented. There are no signs of saturation of the VELO for the reach in $N_{\mathrm{VELO}}^{\mathrm{clusters}}$ of Pbp collisions, so the explanation is likely to be linked to the VELO acceptance as it was explained before.

The behaviour was already observed and discussed when studying the dependence of the reconstruction efficiency with detector occupancy in Sec. 6.2.1 and Fig. 6.3. The feature evidences the need to consider simultaneously the occupancy in different subdetectors to describe properly the fake rates in proton-lead, as it is done with the selected sets of weights. In the case of pp, the tendency is similar for $N_{\mathrm{VELO}}^{\mathrm{clusters}}$ and $N_{\mathrm{SPD}}^{\mathrm{hits}}$.

For the fraction of secondaries, shown in Figs. 6.17c and 6.17b, the dependency is rather flat. The fraction of clone candidates, shown in Figs. 6.17e and 6.17f, is below 0.3% so it is not expected to have an impact in the measurement. In any case, the fraction decreases with the event occupancy in all samples.

Eqs. 6.13, 6.14 and 6.15 can be used to extract each background fraction in simulation. However, these might differ from the fractions of background in data, denoted by f_{fake}, f_{sec} and f_{clone}. It is necessary to have a procedure to infer the differences between data and simulation and to account for them. Since the background contribution to the candidates sample after the selection requirements is reduced, an estimation of its upper and lower limit may be sufficient in the analysis. The fractions in data and simulation are assumed to be proportional,

$$f_\alpha = R_\alpha f_\alpha^{\mathrm{sim}}, \tag{6.16}$$

where $\alpha = fake, sec, clone$ and R_α is the data/simulation discrepancy.

To calculate R_α a high purity *proxy* sample of the α component of the background can be defined by selecting a group of reconstructed long tracks with a high background purity. Then, R_α can be extracted as the data/simulation ratio of this sample. This assumes that the background discrepancy of data and simulation in the proxy sample is similar to the discrepancy in the candidate sample. This hypothesis is subject to

systematic uncertainties discussed in detail in Sec. 7.3. In this way,

$$R_\alpha = \frac{\frac{1}{N_{\text{candidates}}} \sum_i N^{\text{data}}_{\alpha,i}}{\frac{1}{\sum_i w_i N_{\text{cand},i}} \sum_i w_i N^{\text{sim}}_{\alpha,i}}, \tag{6.17}$$

where N_α is the number of long tracks in the proxy sample, w_i is the event occupancy weight of the i-th event of simulation, $N_{\text{cand},i}$ is the number of candidates of the i-th event and $N_{\text{candidates}}$ is the total number of candidates. Note that $N_{\text{candidates}}$ and N_α depend on the (η, p_{T}) bin, resulting in a bin-dependent R_α.

All the events passing the selection (Sec. 5.1) and with at least one reconstructed track within the LHCb acceptance are considered in the proxy samples. With these criteria, every event considered in the weighting procedure is accounted for in Eq. 6.17. The definition of proxy samples for fake tracks and secondary particles is sensitive, and will be studied in the following.

6.4.1 Fake tracks

In Fig. 6.19, the fraction of fake tracks in the candidate sample is shown for Pbp, pPb and pp simulation. These fractions have been obtained using Eq. 6.13. No differences between MD and MU configurations are seen, and therefore both datasets are merged in each of the proton-lead configurations to increase the samples size. The fractions of fake tracks at low p_{T} are higher in Pbp, then pPb and finally pp. At high p_{T}, the three fractions tend to be more similar.

To define a proxy for fake tracks to be used with Eq. 6.17, the sample of reconstructed long tracks within the LHCb acceptance is considered. Additionally, the requirement in `pseudoIP` of Tab. 5.6 is applied to minimise the influence of secondary particles. However, the critical part is to apply a requirement that increases the purity of fake tracks in the proxy sample. Two different strategies are proposed:

1. **Tracks with high ghost probability.** An enriched sample in fake tracks can be obtained using the `GhostP` variable, since the purity of fake tracks strongly increases with ghost probability. With a `GhostP` > 0.5 requirement a fake track purity in the simulation over 97% for the Pbp configuration and over 93% for pPb the configuration is achieved. For the pp configuration the `GhostP` < 0.4 requirement, applied at the reconstruction level, removes high `GhostP` long tracks and limits the applicability of this strategy.

2. **Shared VELO tracks.** The use of shared VELO tracks, which were studied in detail in Sec. 5.3.3, is motivated because they show a much higher proportion of fake tracks than non-shared VELO tracks. This is true throughout the (η, p_{T}) spectra and for Pbp, pPb and pp events, as can be seen in Fig. 5.17 compared to Fig. 5.9. With this strategy higher fake track purities can be achieved with respect to only considering a `GhostP` requirement. A drawback to this strategy is the smaller dataset since only $\sim 1\%$ of the tracks are flagged as shared VELO.

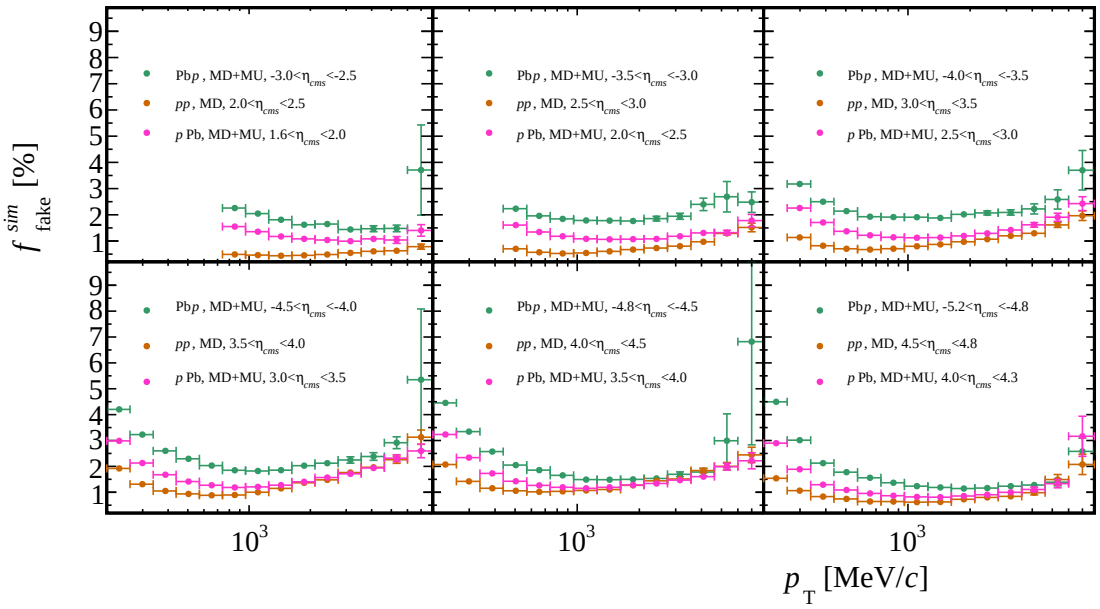


Figure 6.19: Fraction of fake tracks in the candidate sample for the Pbp, pPb and pp simulation, computed with Eq. 6.13. Simulation has been weighted using the weight set 1 for Pbp and pPb and the set 12 for pp. Error bars correspond to statistical uncertainty.

The `GhostP` distribution for tracks meeting the requirements in Tab. 6.4 is shown in Fig. 6.20 to evaluate the two different strategies. The plot is normalised by the number of candidates to replicate the numerator and denominator of Eq. 6.17. No particular (η, p_{T}) bin is chosen, and therefore the plots only show the general picture. The fake track distribution (magenta) is similar for the two strategies, however in the case of the shared VELO tracks they dominate over the signal for values of `GhostP` as low as 0.1, which is not the case for the other proxy that requires a larger `GhostP` cut. Hence, with the shared VELO proxy high purity samples can be achieved with ghost probability requirements similar to those applied to candidates. This is desirable since the data/simulation discrepancy in the fake rate is extrapolated to the candidate sample.

A nominal proxy is defined to compute the value of R_{fake} that will be used to obtain f_{fake}. After an exhaustive study, it was decided to take different choices of the nominal proxy for proton-lead and pp samples due to the presence of the `GhostP` < 0.4 requirement in the pp events. In proton-lead the proxy are non shared VELO tracks with $0.5 <$ `GhostP` < 0.9. In pp events shared VELO tracks with $0.1 <$ `GhostP` < 0.4 are chosen. Exceptionally, the requirement for $6 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$ and $4.5 < \eta < 4.8$ is set to $0 <$ `GhostP` < 0.4 to increase the sample size in the bin.

These criteria are a good compromise between the different aspects that influence the precision of the technique, such as the sample size of the proxy sample in data and simulation for all the (η, p_{T}) bins, the fake track purity and the *proximity* to the candidate

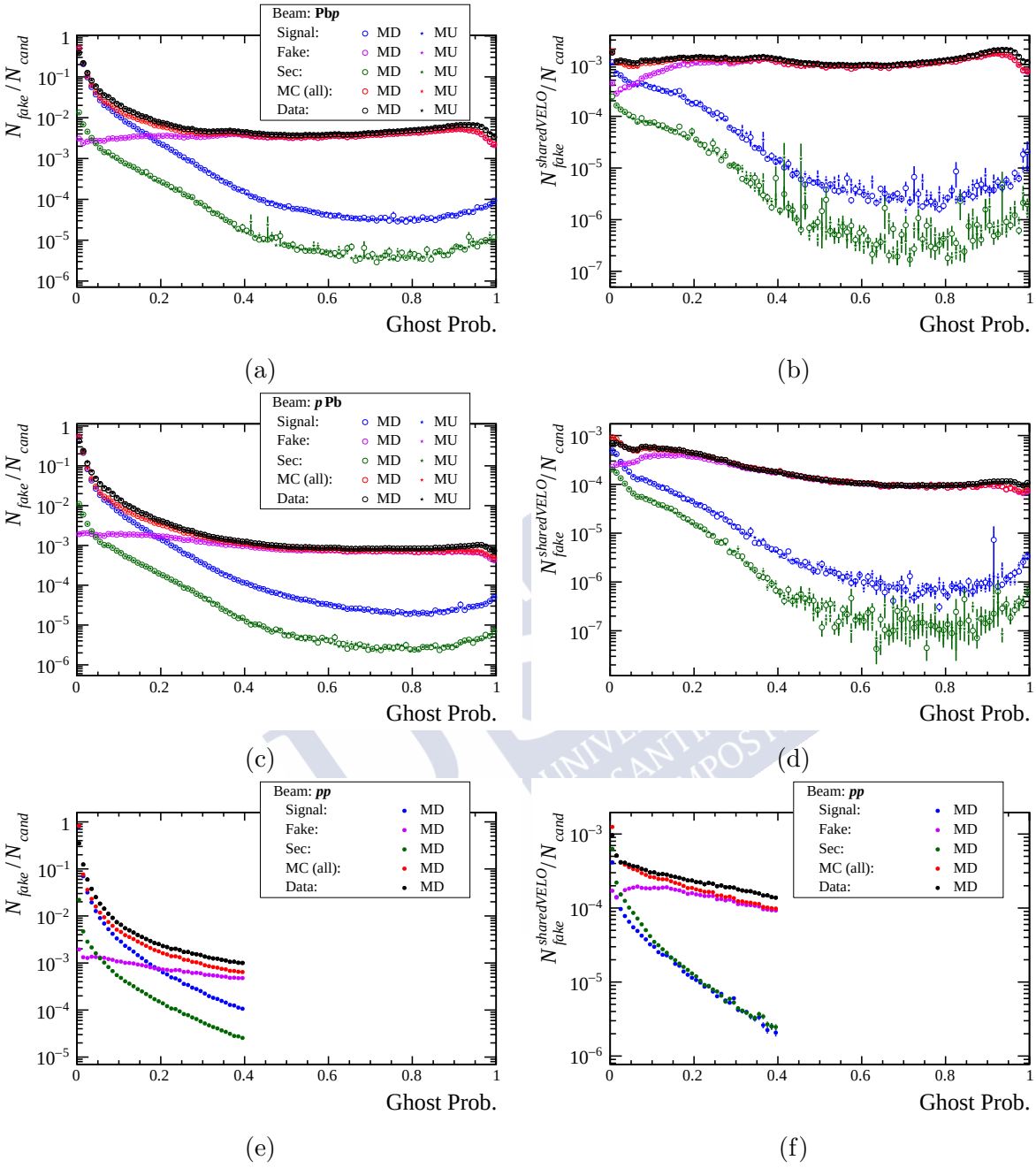


Figure 6.20: Ratio between the number of tracks in the proxy sample and the number of candidates for different bins of `GhostP` in Pbp (top), pPb (middle) and pp (bottom). Left column plots consider every track while in the right column plots contain only shared VELO tracks. Black markers correspond to real data, while the rest of the colours correspond to simulation, being red the sum of all the components. Simulation is reweighted with set 1 for proton-lead and with set 12 for pp.

region. A summary of the requirements applied to define the proxy samples is shown in Tab. 6.4. Alternative proxy definitions are explored to assign the systematic uncertainty to R_{fake}, see Sec. 7.3 for more details.

The fake track purity in the proxy sample seemed to variate significantly with the requirements in `GhostP`. This quantity can be computed only for the simulation sample. In particular, one aspect that complicated the proxy selection was an important variation between the different (η, p_{T}) bins. Fig. 6.21 shows the purity of fake tracks for the *proxy* regions of `GhostP` in simulation. With the chosen requirements, a purity above 75% in every studied bin is achieved, for some cases being almost 100%. This was specially challenging for the *pp* and motivates the use of the shared VELO requirement in that case. Without this strategy, reaching a purity above 50% for the $2 < \eta < 2.5$ bin was found impossible.

The R_{fake} results are shown in Fig. 6.22 for Pb*p*, *p*Pb and *pp* data. The error bars include the statistical and systematic uncertainties summed in quadrature. The uncertainty analysis is explained in detail in Sec. 7.3.

The results indicate that the weighted simulation reproduces the fraction of fake tracks in data for $p_{\text{T}} < 2\,\text{GeV}/c$ in every η bin. Here, the weighting has not a significant impact in *pp*, but it does in proton-lead data specially in Pb*p* due to the strong dependence with detector occupancy of the observable. This was checked by comparing the result obtained with the simulation sample without any weights. For higher p_{T}, the situation changes for each configuration. The *p*Pb trend is to keep the data/simulation agreement, for Pb*p* a small increase of R_{fake} is observed and finally a significant deviation from unity is seen for *pp*. For *pp*, this behaviour is seen for all the tested occupancy weight sets. The behaviour at high p_{T} is hinted in Fig. B.1d, where the different behaviour of fake tracks in data and simulation is evident. In both proton-lead configurations, important deviations from unity are also seen if the simulation is not weighted for the detector occupancy, as can be seen in Fig. B.21 of App. B. This can be a hint to explain the observed behaviour. The high p_{T} region is clearly more challenging due to a much worse signal to background ratio

	pp	*p*Pb	Pb*p*
Track type	Long	Long	Long
η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
p_{T}	$0.2 < p_{\text{T}} < 8\,\text{GeV}/c$	$0.2 < p_{\text{T}} < 8\,\text{GeV}/c$	$0.2 < p_{\text{T}} < 8\,\text{GeV}/c$
p	$p > 2\,\text{GeV}/c$	$p > 2\,\text{GeV}/c$	$p > 2\,\text{GeV}/c$
`pseudoIP`	$< 0.368\,\text{mm}$	$< 0.348\,\text{mm}$	$< 0.348\,\text{mm}$
`GhostP`	$0.1 < \texttt{GhostP} < 0.4^{\dagger}$	$0.5 < \texttt{GhostP} < 0.9$	$0.5 < \texttt{GhostP} < 0.9$
is shared VELO	`True`	`False`	`False`

Table 6.4: List of requirements for the fake track proxy in each configuration. The standard (η, p_{T}) bins are considered.
† For the $6 < p_{\text{T}} < 8\,\text{GeV}/c$ and $4.5 < \eta < 4.8$ bin, the requirement is $0 < \texttt{GhostP} < 0.4$.

and much smaller data and simulation samples. Nevertheless, the computation of R_{fake} is robust and the assigned systematic uncertainty covers the variation observed in different tests.

Using these values for R_{fake}, the fractions of fake tracks in data can be computed and are shown in Fig. 6.23. The fraction of fake tracks is found to be below 5% in most bins, which is achieved thanks to the strong requirement in `GhostP` applied to the candidates.

6.4.2 Secondary particles

Before studying its proxy definition, the composition of secondary particles from the different sources listed in Sec. 5.3.2, needs to be studied. This can be done using information of the corresponding MCparticle of every candidate flagged as non-prompt. Useful details are the particle ID, the mother ID and the origin vertex type (`OV_type`). This last property indicates how the particle was produced in simulation. For instance, the `OV_type` is 1 if the MCparticle originates in the primary interaction, 2 from a particle decay, 100 from a hadronic interaction with the detector material and 102 from γ conversion.

Fig. 6.24 shows the background composition of secondary particles with respect to η and p_{T} for pPb, Pbp and pp simulation. The dependency with respect to p is shown in Fig. 6.25. The different contributions are separated into the following categories motivated by the proxy definitions discussed later:

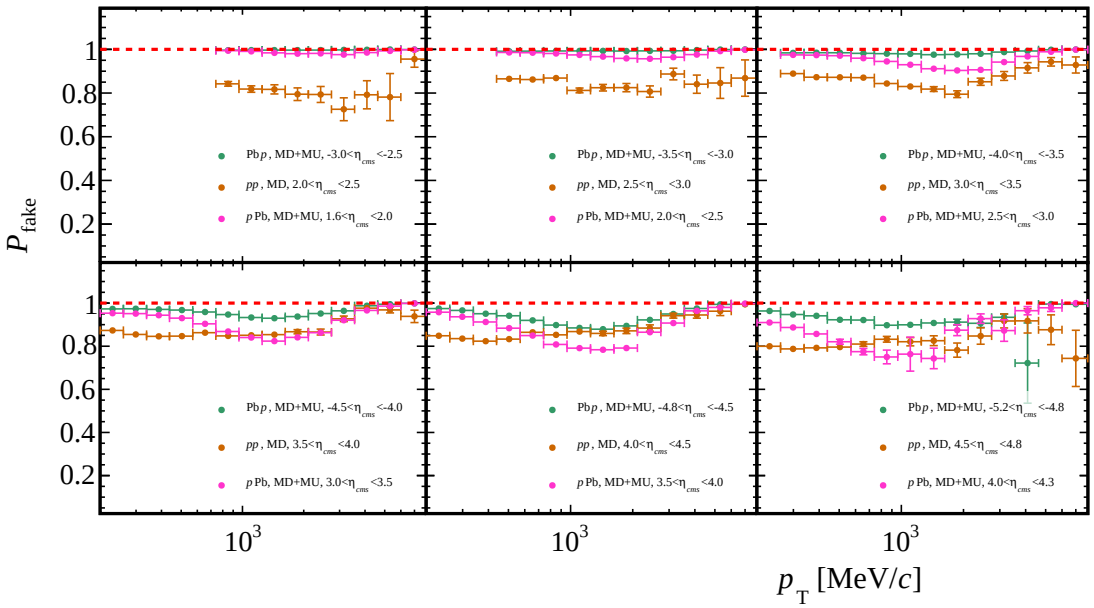


Figure 6.21: Fake track purity for the nominal proxies as defined in Tab. 6.4 for pPb, Pbp and pp simulation samples. Error bars include statistical uncertainties.

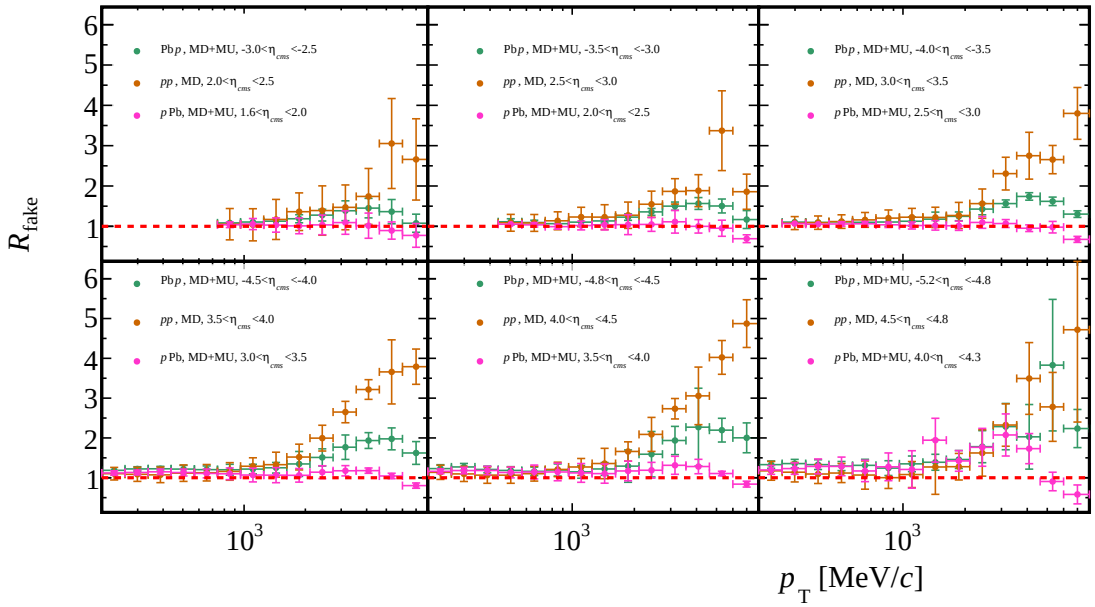


Figure 6.22: Results for the discrepancy data over simulation R_{fake} for fake tracks in Pbp, pPb and pp data sets. Simulation is weighted with set 1 for proton-led and set 12 for pp. The values have been obtained using Eq. 6.17 and with the fake track proxy requirements of Tab. 6.4. Error bars include statistical and systematic uncertainty added quadratically.

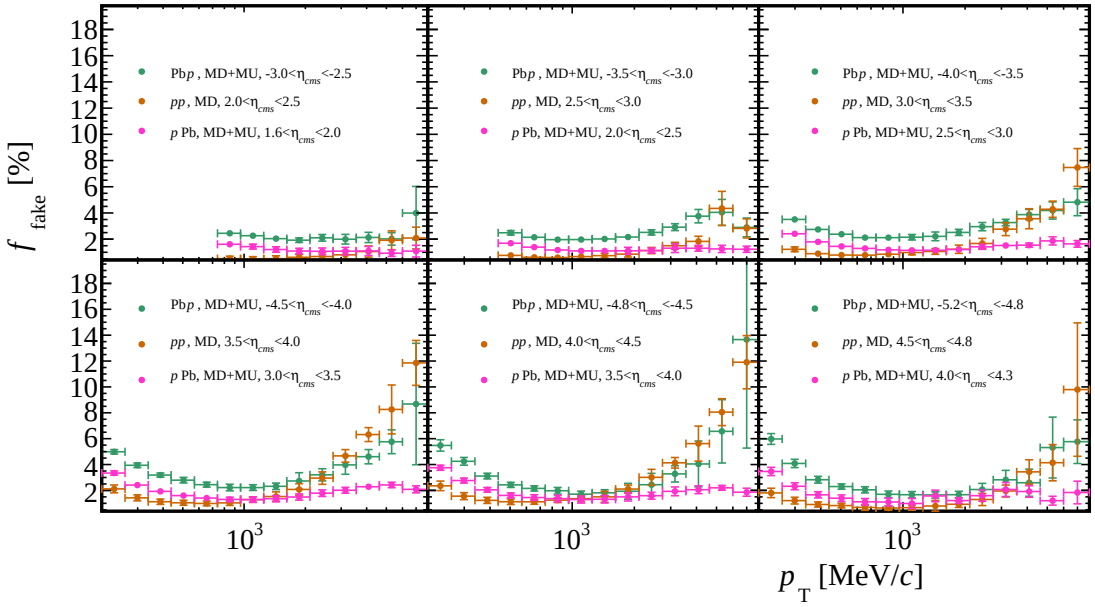


Figure 6.23: Fraction of fake tracks in the distribution of candidates for the Pbp, pPb and pp data samples, obtained using Eq. 6.16.

- $X \to \mu$: muons from particle decays.
- $\gamma \to e$: converted photons.
- *Had*: particles produced in hadronic interactions with the detector material.
- *Decays*. This category contains different contributions.
  - $\Lambda \to h$; $K^0_S \to h$; $\Sigma^{\pm} \to h$; $\Xi, \Omega \to h$: hadrons originated in decays of prompt particles.
  - $X \to e$: electrons from particle decays.
  - *Other decays*: other non-prompt particles not included in the previous categories.

In fact, the last two categories (*Had* and *Decays*) will be jointly described by a single proxy as afterwards discussed.

From Fig. 6.24, the converted γ background dominates, specially at low p_{T}. The second most abundant background is decays of hadrons, of which K^0_S and Λ are the most relevant. Contribution from hadronic interactions is relatively important in Pbp, mostly at high p_{T}. The $X \to \mu$ contribution is noticeable at low p_{T}.

The fraction of secondary particle candidates can be factorised as:

$$f^{\mathrm{sim}}_{\mathrm{sec}} = f^{\mathrm{sim}}_{\gamma} + f^{\mathrm{sim}}_{X\to\mu} + (f^{\mathrm{sim}}_{\mathrm{had}} + f^{\mathrm{sim}}_{\mathrm{decays}}), \tag{6.18}$$

where, with the two terms within brackets, the combined consideration of these two contributions is anticipated. In Figs. 6.26 and 6.27, the fraction of secondary particles in the candidate distribution of the Pbp, pPb and pp simulation samples is shown, with the contribution from each category. As no differences between MD and MU are seen in Fig. 6.24, both datasets are merged. The landscape is similar in the three configurations: the converted γ contribution is the largest but is only important at low p_{T}, the contribution from decays is more uniform along p_{T} and the contribution from hadronic interactions seems to play a role only at high p_{T} for the Pbp sample. The contribution from decays to muons is very small for all bins.

The next step is to identify variables that allow to isolate the different background contributions from secondary particles to apply Eq. 6.17. Ideally, one should be able to isolate one secondary component from other contributions in the long track spectra (prompt charged particles, fake tracks or other secondaries). Three variables are used to construct the proxy samples:

1. DLL$_{e\pi}$. The RICH PID variable for electron-pion discrimination, allows to isolate background from converted γ.
2. DLL$_{\mu\pi}$. The RICH PID variable for muon-pion discrimination, isolates background from decays into muons.

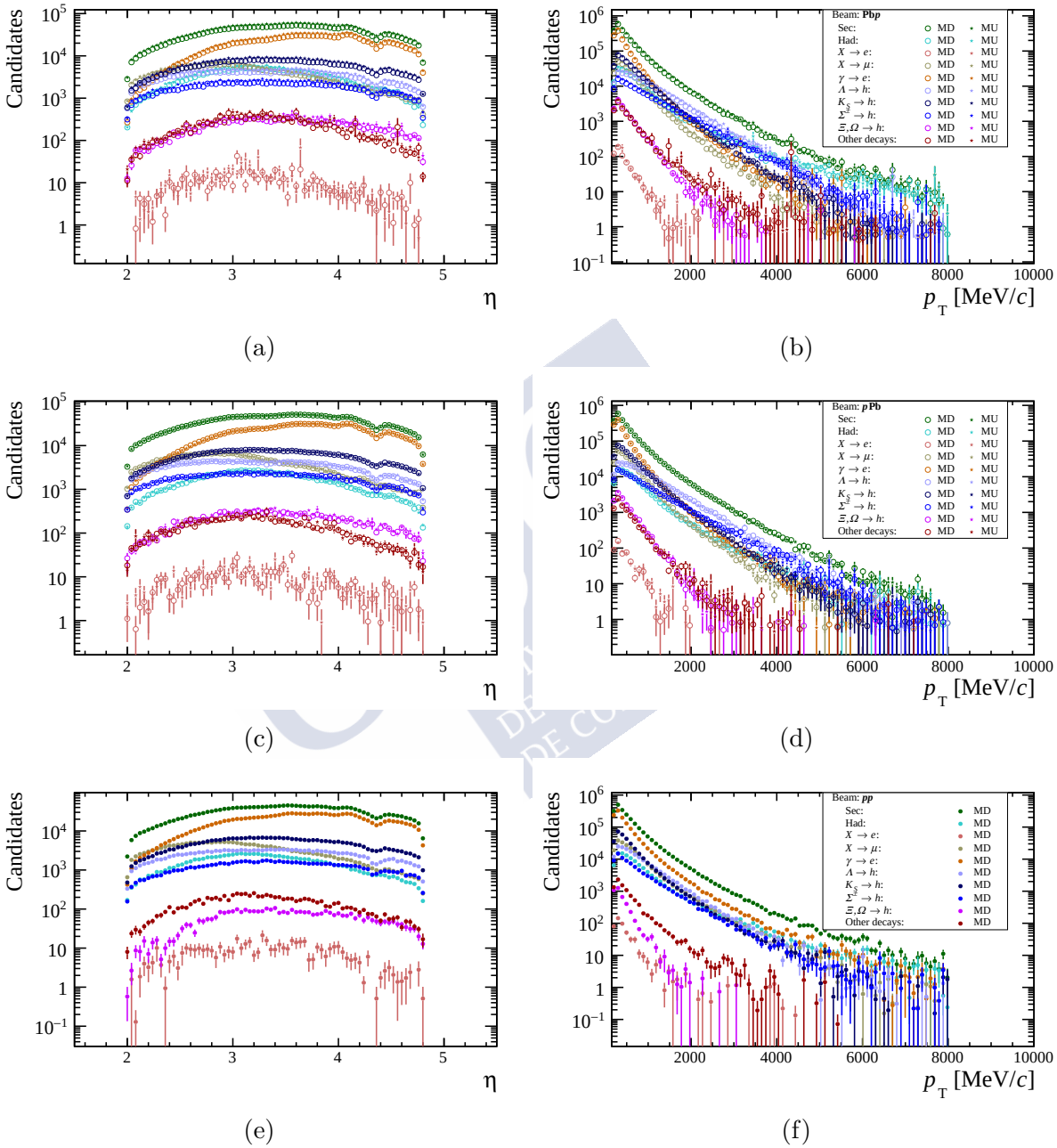


Figure 6.24: Composition of the background from secondary particles with respect to η (left) and p_T (right) for Pbp (top), pPb (middle) and pp (bottom).

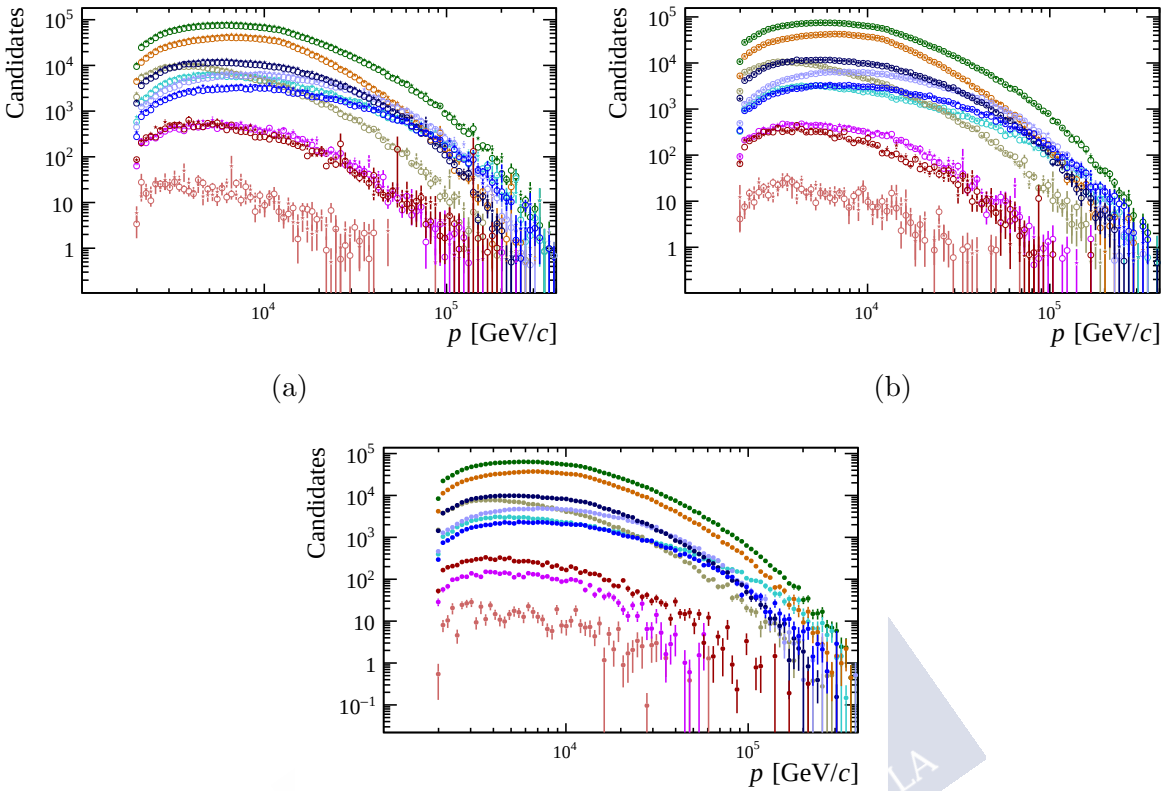
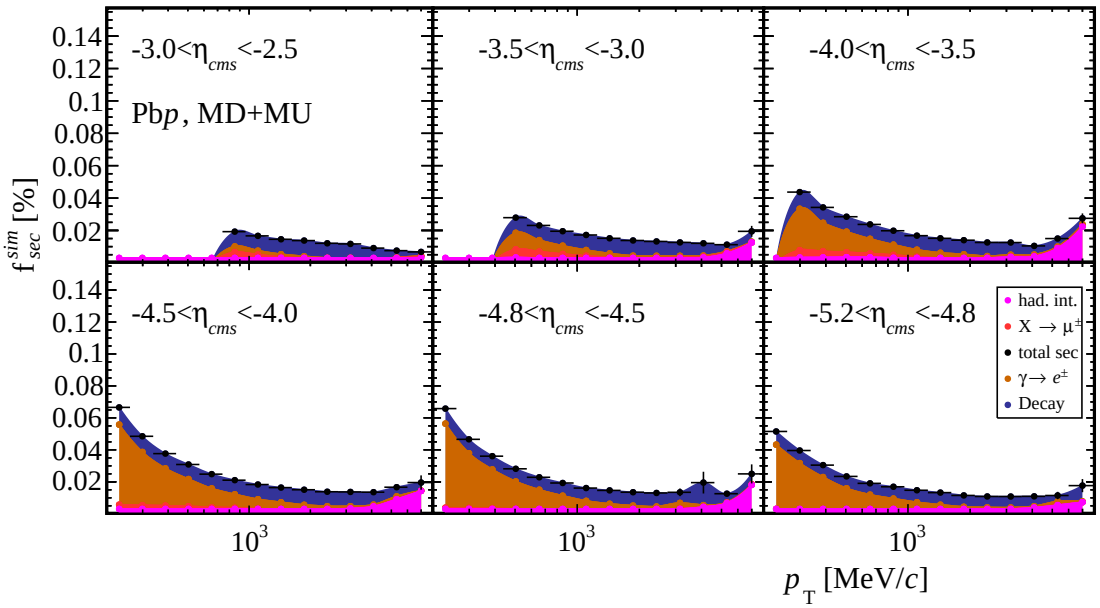


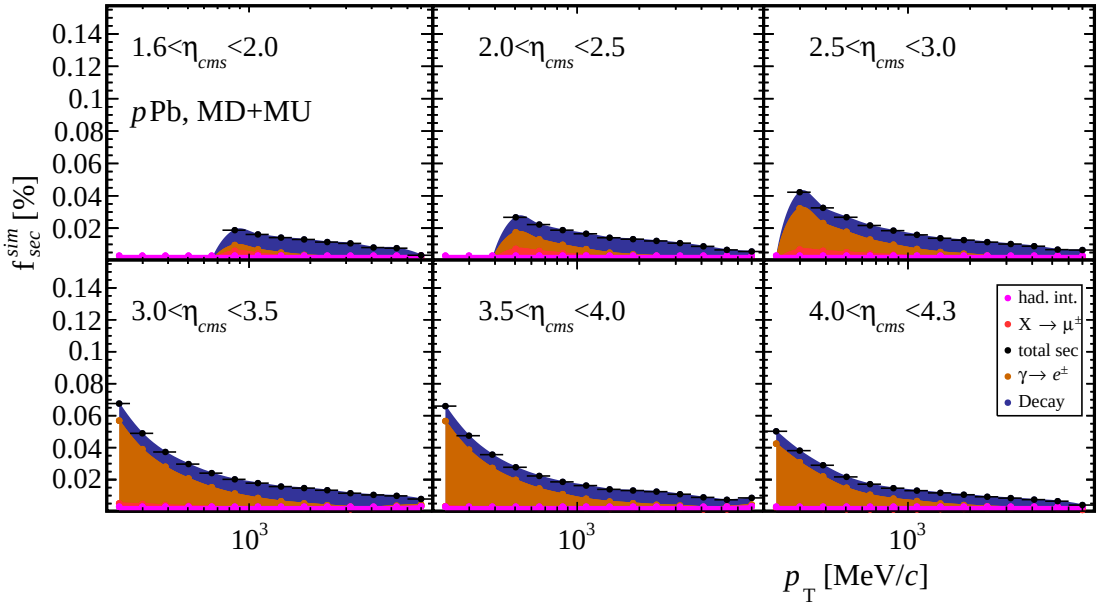
Figure 6.25: Composition of the background from secondary particles with respect to p (left top) for Pbp, pPb (right top) and pp (bottom). The legend is shared with Fig. 6.24.

3. `pseudoIP`. Tracks with high `pseudoIP` are generally produced far from the interaction point. Thus, a `pseudoIP` cut can be applied to separate particles from hadronic interactions and particle decays.

The distributions of these variables in data and simulation are presented in Figs. 6.28 and 6.29. The main contributions in simulation are presented as well to give an idea of the discrimination power of each variable. The plots are normalised to the number of candidates in data and simulation respectively following the normalisation of Eq. 6.17. The distributions include long tracks with some clean-up requirements which are applied in order to increase the purity of the background of interest for each case. These requirements are indicated in Tab. 6.5. Both $\mathrm{DLL}_{e\pi}$ and $\mathrm{DLL}_{\mu\pi}$ can be used to isolate the converted γ and $X \to \mu$ contributions. On the other hand, `pseudoIP` separates the contributions from hadronic interactions and other decays from the rest. Since these two contributions cannot be easily separated, they are integrated in a single contribution called *Had&decays*. The



(a)



(b)

Figure 6.26: Composition of the background from secondary particles in Pbp (Fig. 6.26a) and pPb (Fig. 6.26b) simulation separated by category.

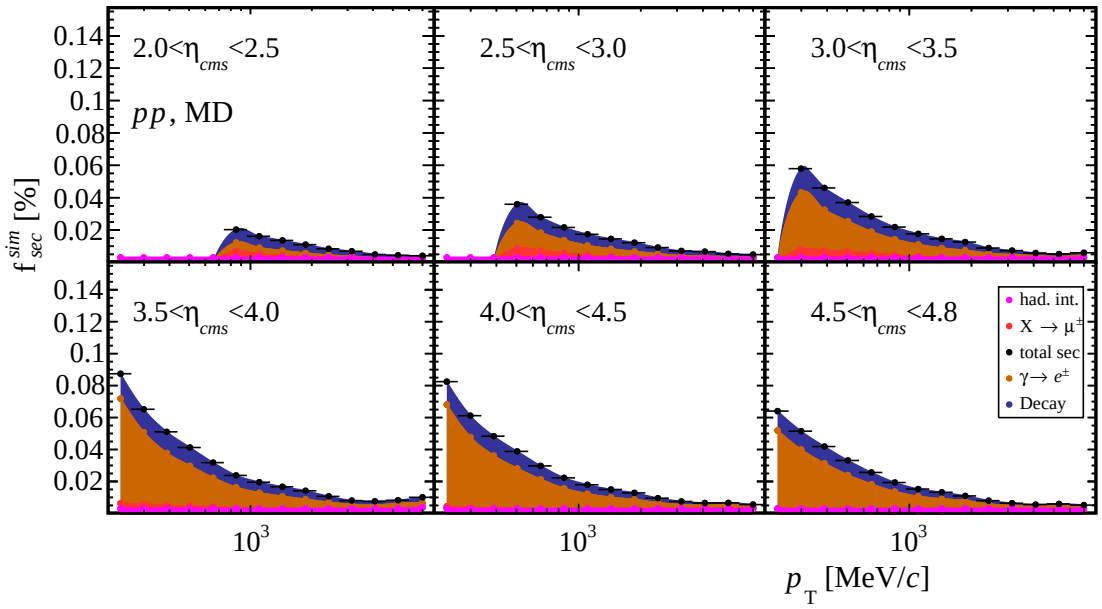


Figure 6.27: Composition of the background from secondary particles in pp simulation separated by category.

equation for the fraction of secondary particles becomes

$$f_{\text{sec}} = R_{\gamma} f_{\gamma}^{\text{sim}} + R_{\text{had\&decays}} f_{\text{had\&decays}}^{\text{sim}} + R_{X\to\mu} f_{X\to\mu}^{\text{sim}}. \tag{6.19}$$

Where the R_{γ}, $R_{\text{had\&decays}}$ and $R_{X\to\mu}$ coefficients are the data/simulation discrepancies that are measured with Eq. 6.17. Each of the three coefficients are evaluated from samples where the requirements from Tab. 6.5 have been applied. Since the distributions presented in Figs. 6.28 and 6.29 depend on the (η, p_{T}) bin, different requirements where tested to optimise each sample size and its purity in simulation. All the tested requirements are indicated within brackets in Tab. 6.5. This strategy is specially needed for $R_{\text{had\&decays}}$ because the `pseudoIP` distribution shows a strong dependence with p, as shown in Fig. 6.30. As p increases, the `pseudoIP` distribution decreases more sharply, since particles with higher momentum tend to be more aligned to the PV. The behaviour is similar in pPb and pp.

The $\text{DLL}_{e\pi}$ distribution showed in Figs. 6.28 and 6.29 is quite stable for the different (η, p_{T}) bins. However, the $4.0 < \eta < 4.5$ interval is an exception. Fig. 6.31 shows how the converted γ tail drops more rapidly than for the neighbouring bins, specially at low p_{T}. This behaviour means a worse performance of the $e - \pi$ separation, and is probably related to the presence of the conical beam pipe inside RICH at 4.38, already mentioned in Sec. 5.3.1. This needs to be considered to set the requirement in $\text{DLL}_{e\pi}$ for the proxy sample.

The next step is to choose a requirement among the ones presented in Tab. 6.5. The

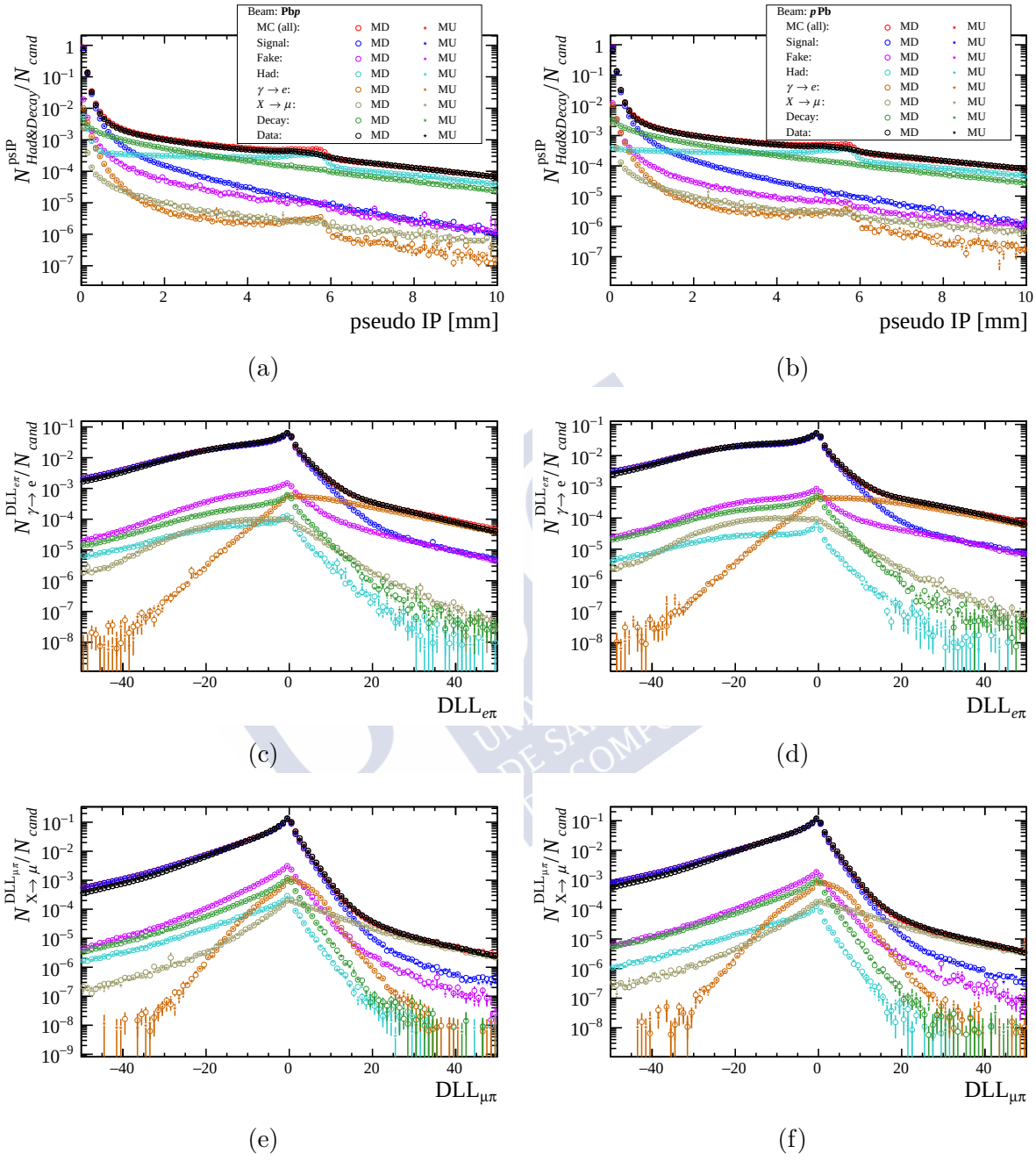


Figure 6.28: `pseudoIP`, $\mathrm{DLL}_{e\pi}$ and $\mathrm{DLL}_{\mu\pi}$ distributions for Pbp (left) and pPb (right). Black markers correspond to real data, while the rest of the colours correspond to simulation, which is weighted with set 1 for the case of Pbp and pPb. Requirements from Tab. 6.5 are applied and each sample is normalised to the number of candidates.

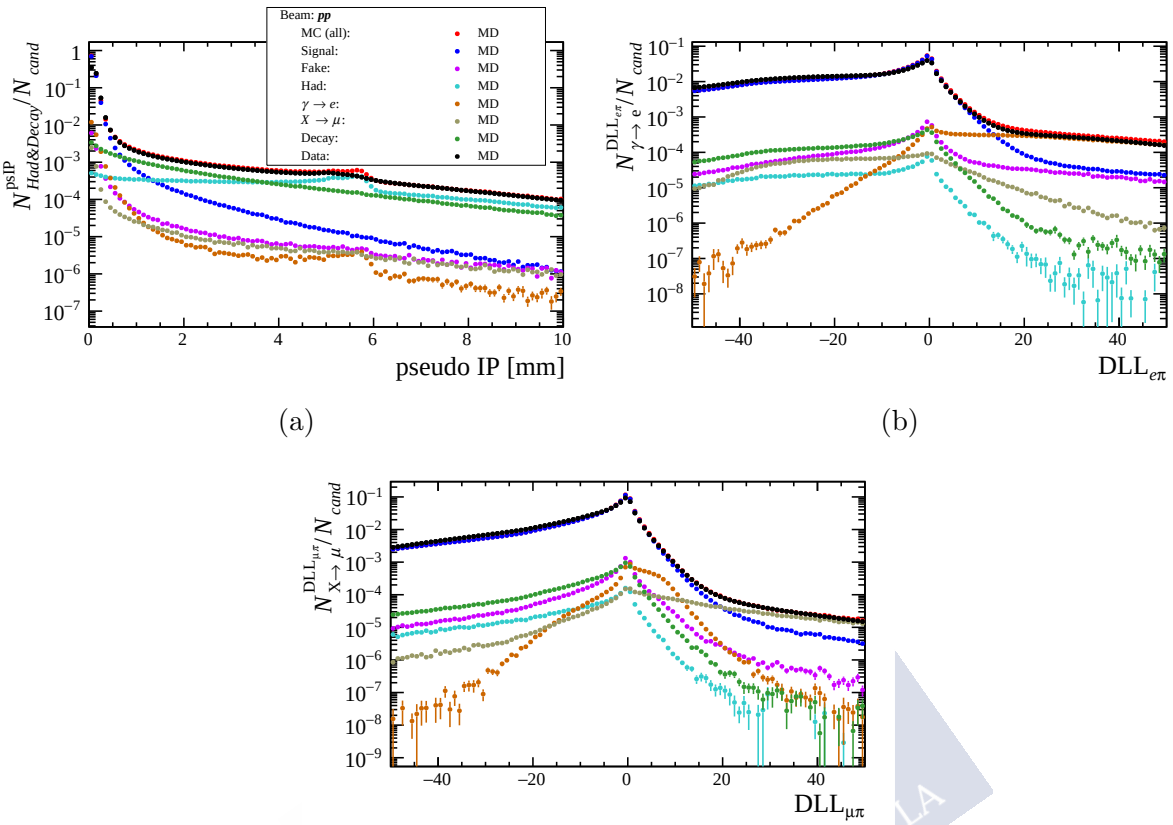


Figure 6.29: `pseudoIP`, $\mathrm{DLL}_{e\pi}$ and $\mathrm{DLL}_{\mu\pi}$ distributions for pp. Black markers correspond to real data, while the rest of the colours correspond to simulation, which is weighted with set 12 for pp. Requirements from Tab. 6.5 are applied and each sample is normalised to the number of candidates.

criteria are the purity of the background component of interest in the resulting proxy and the statistical uncertainty of data and simulation samples. These are the ingredients that influence the precision of R in Eq. 6.17. These criteria are also considered in Sec. 7.3.2 to obtain the systematic uncertainty.

For $R_{\mathrm{had\&decays}}$ the quantity

$$E = \sqrt{u^2_{\mathrm{data}} + u^2_{\mathrm{sim}} + u^2_{\mathrm{purity}}}, \tag{6.20}$$

where $u_{\mathrm{data,sim}}$ is the statistical uncertainty on $N^{\mathrm{data,sim}}_{\mathrm{had\&decay}}$ and $u_{\mathrm{purity}} = 1 - P_{\mathrm{had\&decay}}$. The value of E is computed for each `pseudoIP` requirement in Tab. 6.5. The value for which E is lower is taken to define the proxy.

A similar procedure is followed for R_γ and $R_{X\to\mu}$, however the expression is slightly

	pp	pPb	Pbp
	Common requirements for R_γ, $R_{X\to\mu}$, $R_{\text{had\&decays}}$		
Track type	Long	Long	Long
η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
p_{T}	$0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$	$0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$	$0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$
p	$p > 2\,\mathrm{GeV}/c$	$p > 2\,\mathrm{GeV}/c$	$p > 2\,\mathrm{GeV}/c$
`GhostP`	`GhostP` < 0.078	`GhostP` < 0.103	`GhostP` < 0.109
is shared VELO	`False`	`False`	`False`
	R_γ requirements		
`pseudoIP`	$< 0.368\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$
$\mathrm{DLL}_{e\pi}$	$> [15, 20, 25, 30, 35, 40, 45, 50, 55, 60]$		
	$R_{X\to\mu}$ requirements		
`pseudoIP`	$< 0.368\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$
$\mathrm{DLL}_{e\pi}$	< 15	< 15	< 15
$\mathrm{DLL}_{\mu\pi}$	$> [15, 20, 25, 30, 35, 40, 45]$		
	$R_{\text{had\&decays}}$ requirements		
`pseudoIP`	$> [0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]\,\mathrm{mm}$		

Table 6.5: Requirements considered to define the proxies for R_γ, $R_{X\to\mu}$ and $R_{\text{had\&decays}}$ in each configuration.

changed to

$$E' = \sqrt{u_{stat}^2 + u_{purity}^2}, \tag{6.21}$$

where u_{stat} is directly the uncertainty on R which results from propagating the statistical uncertainties of $N_{\gamma,X\to\mu}^{\text{data,sim}}$ and $N_{\text{cand}}^{\text{data,sim}}$. The change with respect to Eq. 6.20 is motivated by the different behaviour of the $4.0 < \eta < 4.5$ interval as shown in Fig. 6.31.

Once the proxy samples have been defined, Eq. 6.17 can be applied and the values of R can be obtained. However, the purity of the background of interest is to be checked first to validate the procedure. Such purities are shown in Figs. 6.32, 6.33 and 6.34. For most bins background purity rises above 75%. In bins that lack a purity value none of the requirements have entries in N_{bkg}^{sim} so R cannot be obtained. This occurs for high p_{T} bins of P_γ and most bins of P_μ. However, comparing these figures with Figs. 6.26 and 6.27 it is clear than these empty bins correspond to kinematic regions where the corresponding background has no relevance. $P_{\text{had\&decays}}$ is limited for high p_{T} as a result of the behaviour of the `pseudoIP` distribution for high p (Fig. 6.30).

The results for R_γ, $R_{X\to\mu}$ and $R_{\text{had\&decays}}$ are presented in Figs. 6.35, 6.36 and 6.37. For bins without a valid proxy, the value $R = 1 \pm 0.5$ is set. The systematic uncertainty

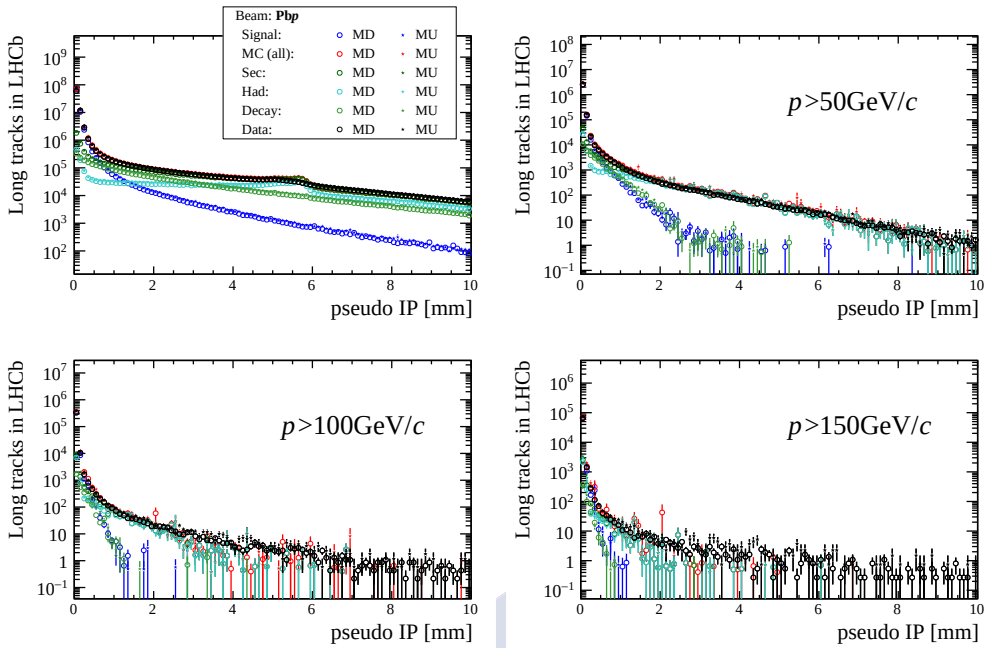


Figure 6.30: Distribution of `pseudoIP` for different p requirements for the Pbp configuration.

is computed as explained in Sec. 7.3.2 and is included in the error bars.

R_γ (Fig. 6.35) has a value around unity for pPb and Pbp, meaning that background in simulation reproduces the data. For pp, simulation overestimates this background component. The reason for this is that the pp simulation was generated with spillover, which increases the number of converted photons. However, pp data, although taken with 25 ns of bunch spacing, comes from the `NOBIAS` stream that includes only events from leading bunch crossings which are free from spillover. This hypothesis has been confirmed with a private simulation excluding the spillover. As the background from simulation is data-driven corrected, there is no effect in the final measurement. $R_{X\to\mu}$ and $R_{\text{had\&decays}}$ (Figs. 6.36 and 6.37) are closer to unity within the given uncertainty for all beam configurations. $R_{\text{had\&decays}}$ for pp is larger than for Pbp and pPb around from 1 to 5 GeV/c of p_{T}, which is expected given the underestimation in Pythia of the K^0_{S} and Λ yields studied in Sec. 5.4.3.

In data, the fraction of secondary particles in the candidate distribution can be obtained applying Eq. 6.19. The results are presented in Figs. 6.39 and 6.38 for pp, pPb and Pbp. A similar contribution from converted γs is found for the three configurations, except for the $4.0 < \eta < 4.5$ interval where the peak at low p_{T} in proton-lead is higher than for pp. At higher p_{T}, background is larger in Pbp than in pPb and pp, mostly due to the contribution from hadronic interactions.

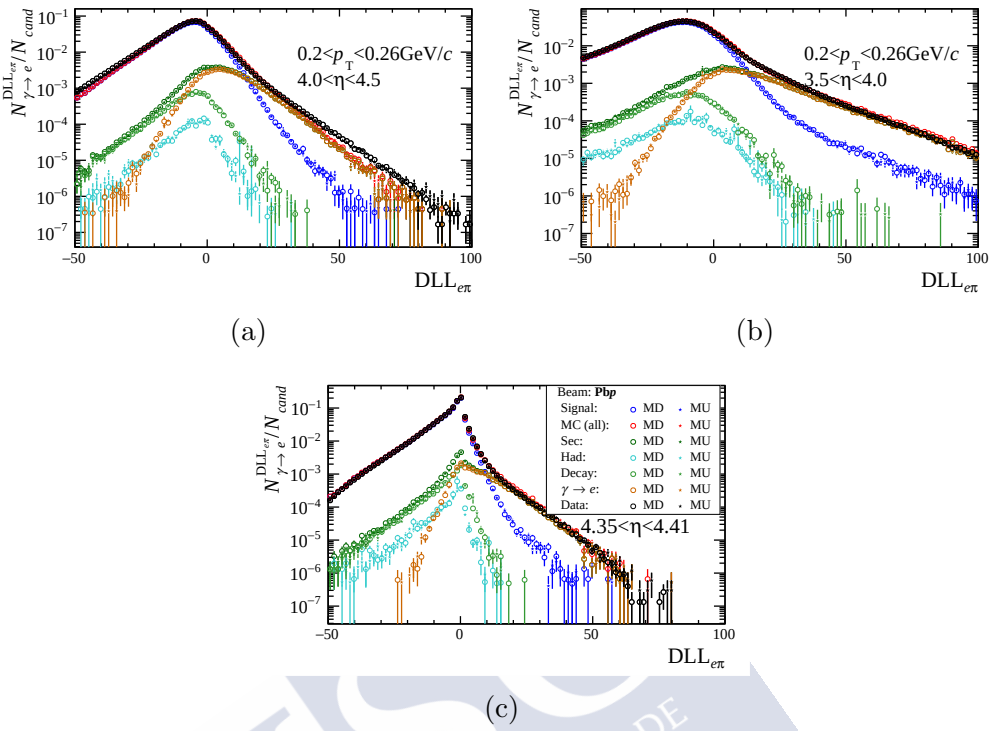


Figure 6.31: Distribution of $\mathrm{DLL}_{e\pi}$ in Pbp configuration for different kinematic ranges. Fig. 6.31a shows the lowest p_{T} bin where the conical beam pipe is present, and Fig. 6.31b shows the neighbour η bin for reference. Fig. 6.31c is a close up on $\eta = 4.38$, where the structure is located.

6.4.3 Summary

After the fake and secondary tracks have been determined, the last background component to be considered are clone tracks. Their fraction ($f^{\mathrm{sim}}_{\mathrm{clone}}$) is obtained with simulation and shown in Fig. 6.40. Since $f^{\mathrm{sim}}_{\mathrm{clone}} < 0.5\%$, this result is directly considered to compute the background fraction in data and no correction with data is applied.

The prompt charge particle purity P in data is obtained applying the following equation:

$$P = 1 - (R_{\mathrm{fake}} f^{\mathrm{sim}}_{\mathrm{fake}} + R_{\gamma} f^{\mathrm{sim}}_{\gamma} + R_{X\to\mu} f^{\mathrm{sim}}_{X\to\mu} + R_{\mathrm{had\&decays}} f^{\mathrm{sim}}_{\mathrm{had\&decays}} + f^{\mathrm{sim}}_{\mathrm{clone}}). \quad (6.22)$$

The final result for P is shown in Fig. 6.41. The error bars include the systematic uncertainties, explained in detail in Sec. 7.3.

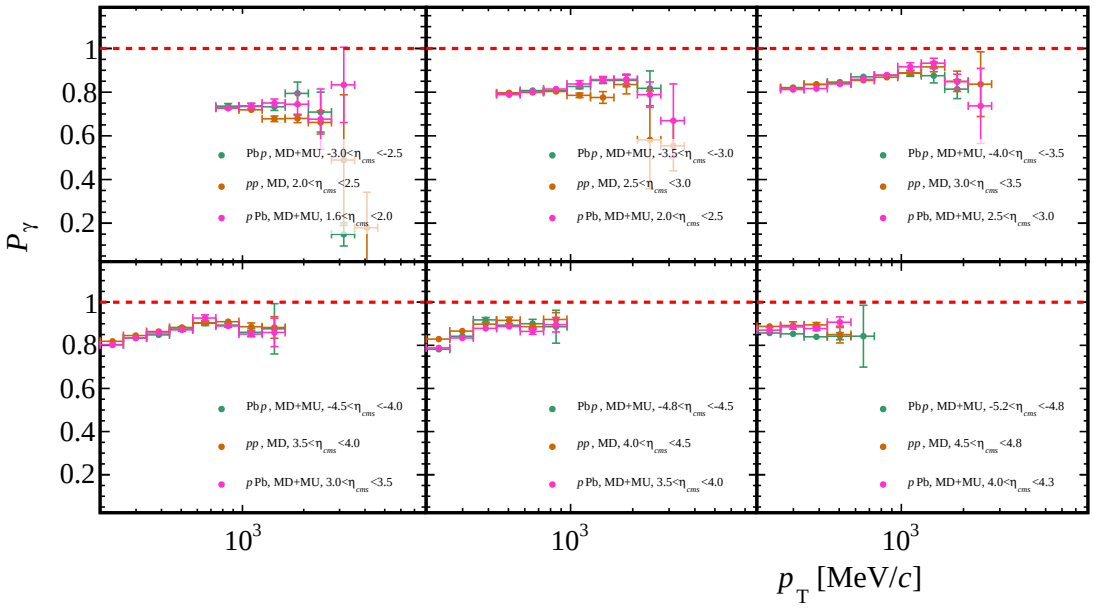


Figure 6.32: Purity of the $\gamma \to e$ contribution in the proxy sample for R_γ in Pbp, pPb and pp collisions in simulation, considering the criteria of Tab. 6.5.

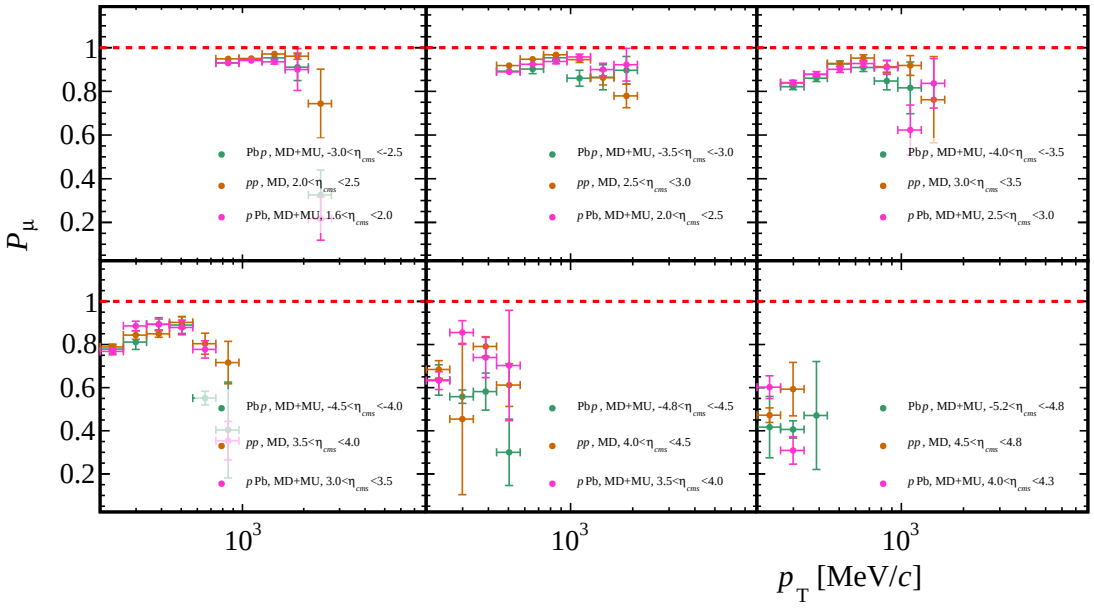


Figure 6.33: Purity of the $X \to \mu$ contribution in the proxy sample for $R_{X\to\mu}$ in Pbp, pPb and pp collisions in simulation, considering the criteria of Tab. 6.5.

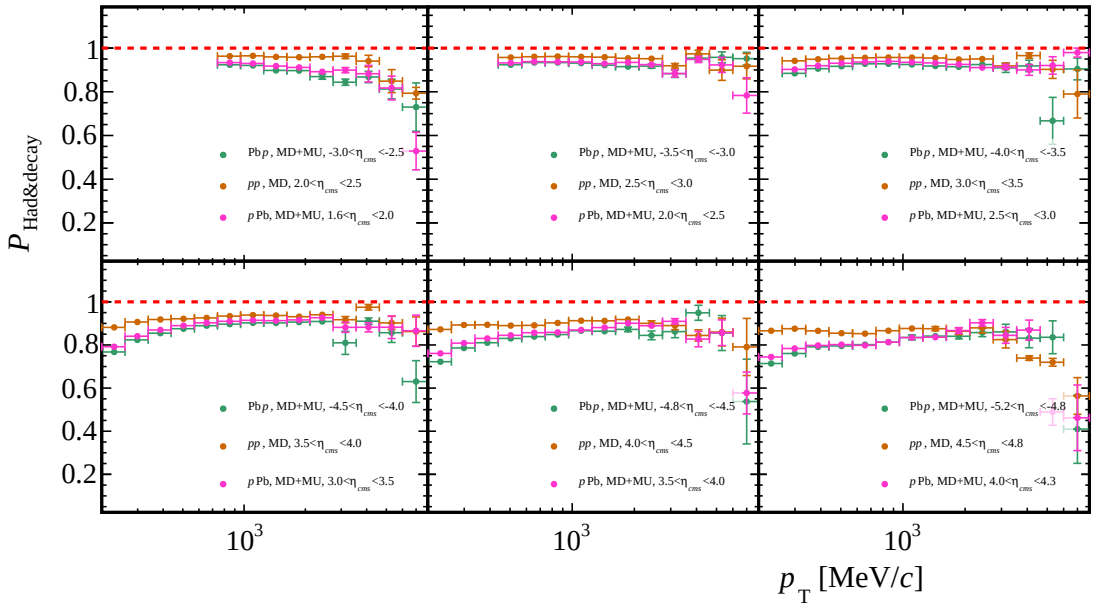


Figure 6.34: Purity of the Had&Decays contribution in the proxy sample for $R_{\text{had\&decays}}$ in Pbp, pPb and pp collisions in simulation, considering the criteria of Tab. 6.5.

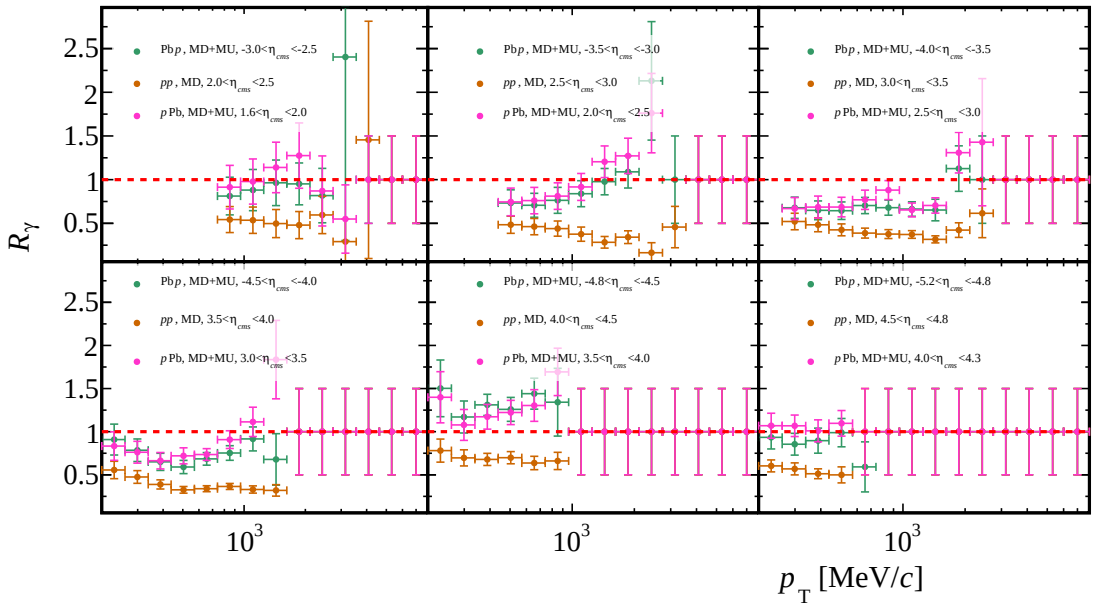


Figure 6.35: Results for R_γ in Pbp, pPb and pp collisions. The values have been obtained using Eq. 6.17 and considering the criteria of Tab. 6.5.

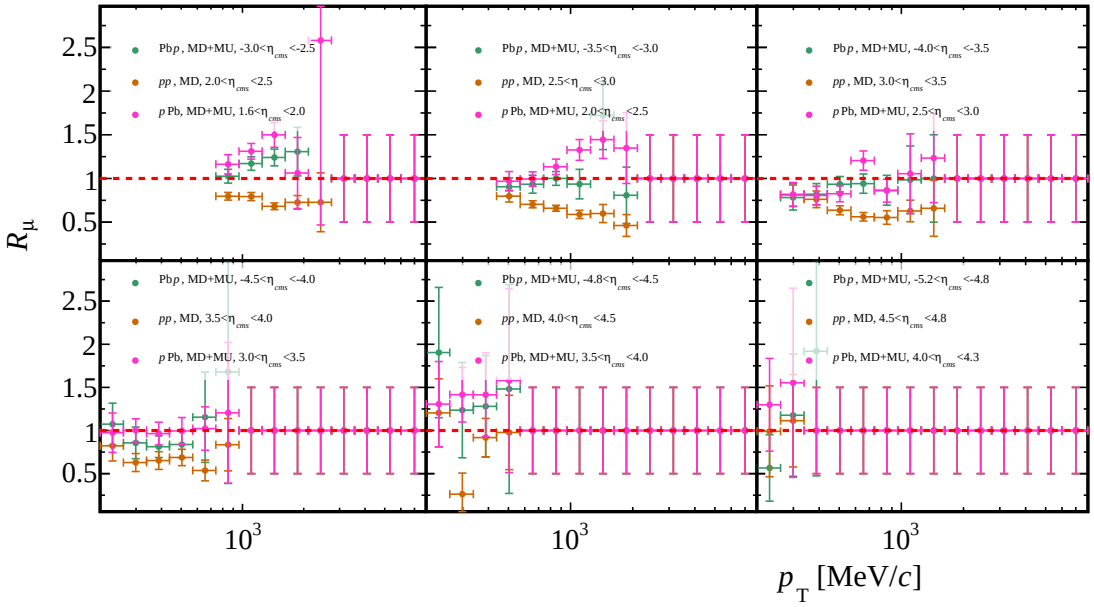


Figure 6.36: Results for $R_{X\to\mu}$ in Pbp, pPb and pp collisions. The values have been obtained using Eq. 6.17 and considering the criteria of Tab. 6.5.

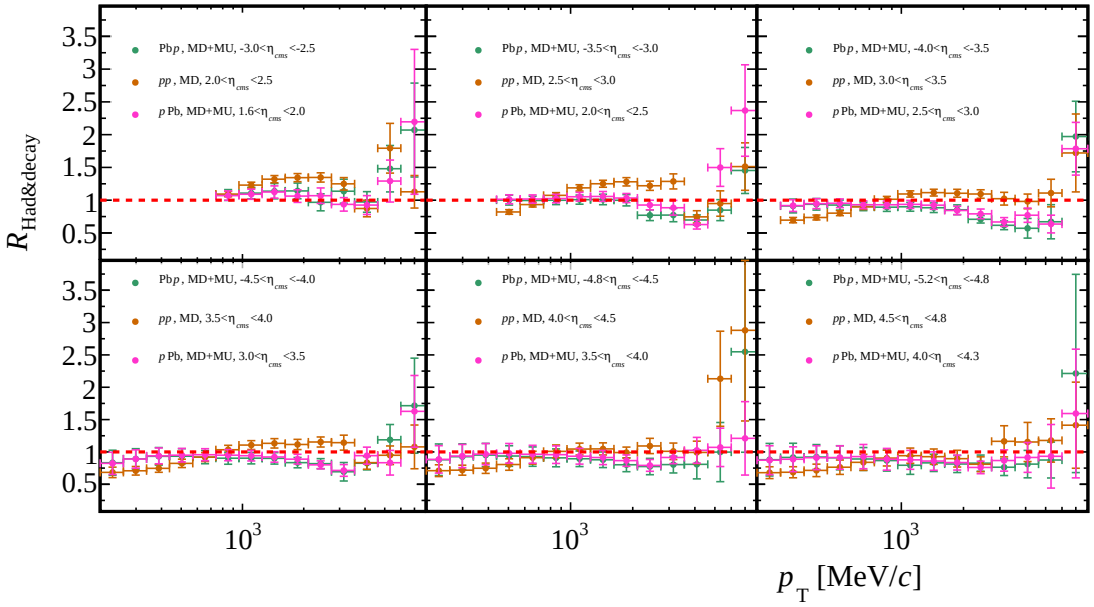


Figure 6.37: Results for $R_{\text{had\&decays}}$ in Pbp, pPb and pp collisions. The values have been obtained using Eq. 6.17 and considering the criteria of Tab. 6.5.

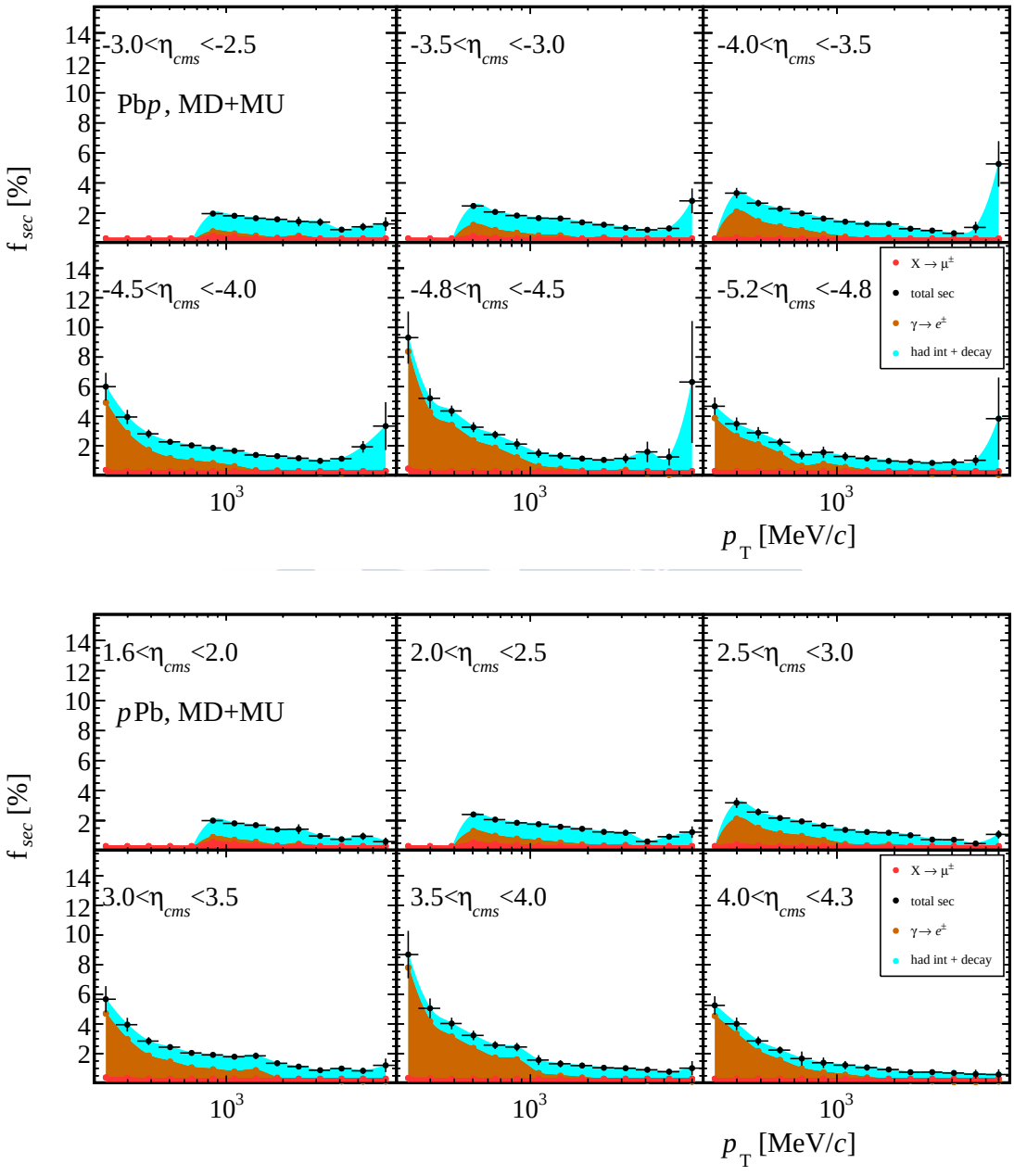


Figure 6.38: Fraction of secondary particles in the distribution of candidates for the Pbp (top) and pPb (bottom) data samples, obtained using Eq. 6.16.

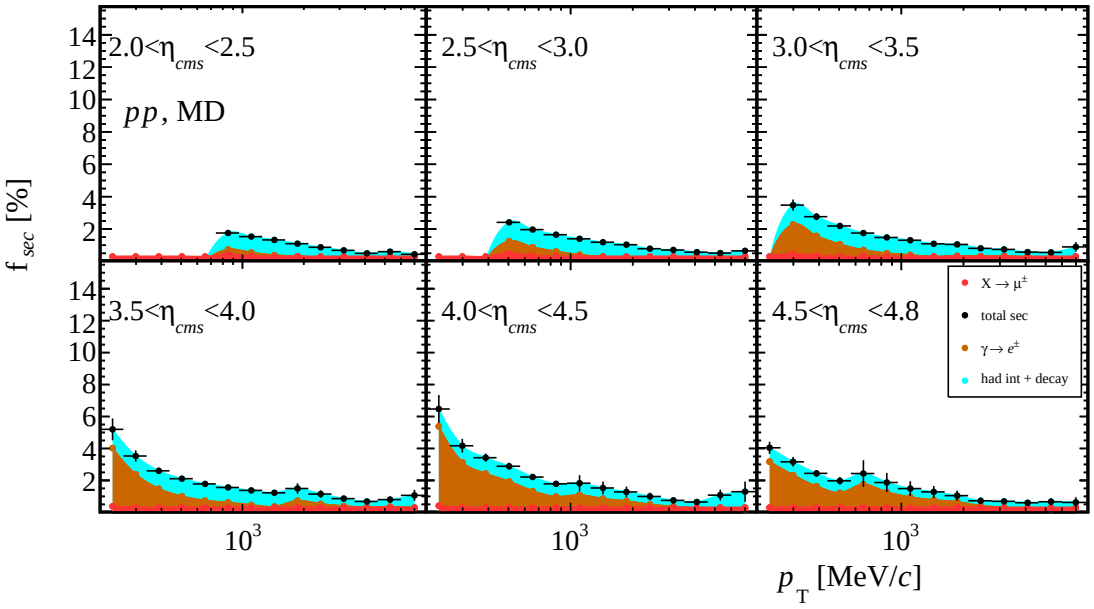


Figure 6.39: Fraction of secondary particles in the distribution of candidates for the *pp* data samples, obtained using Eq. 6.16.

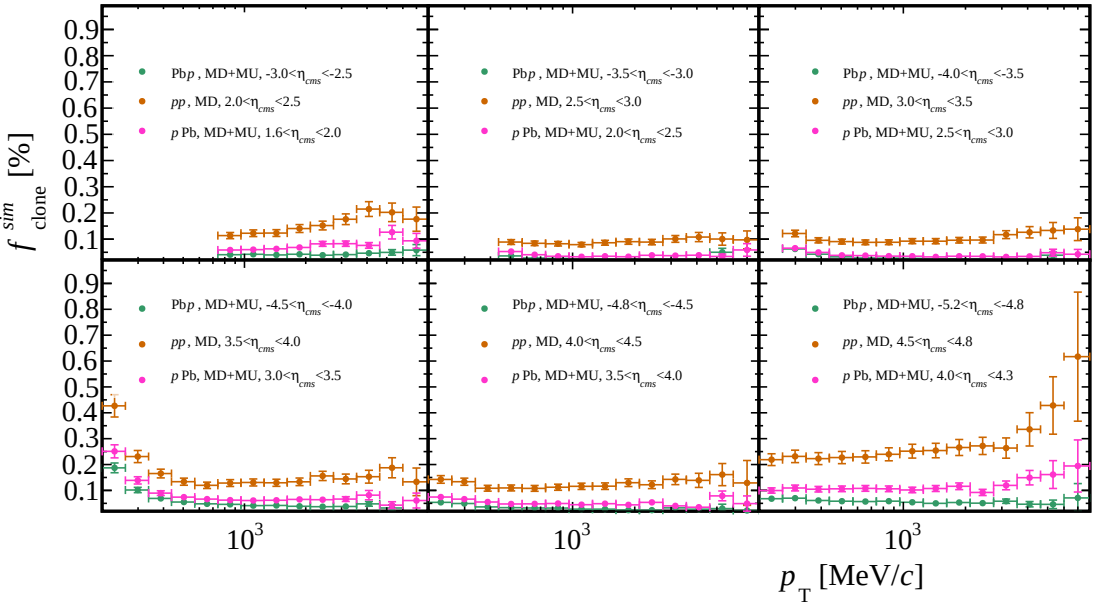


Figure 6.40: Fraction of clone tracks in the distribution of candidates for the Pb*p*, *p*Pb and *pp* simulation samples, computed with Eq. 6.15.

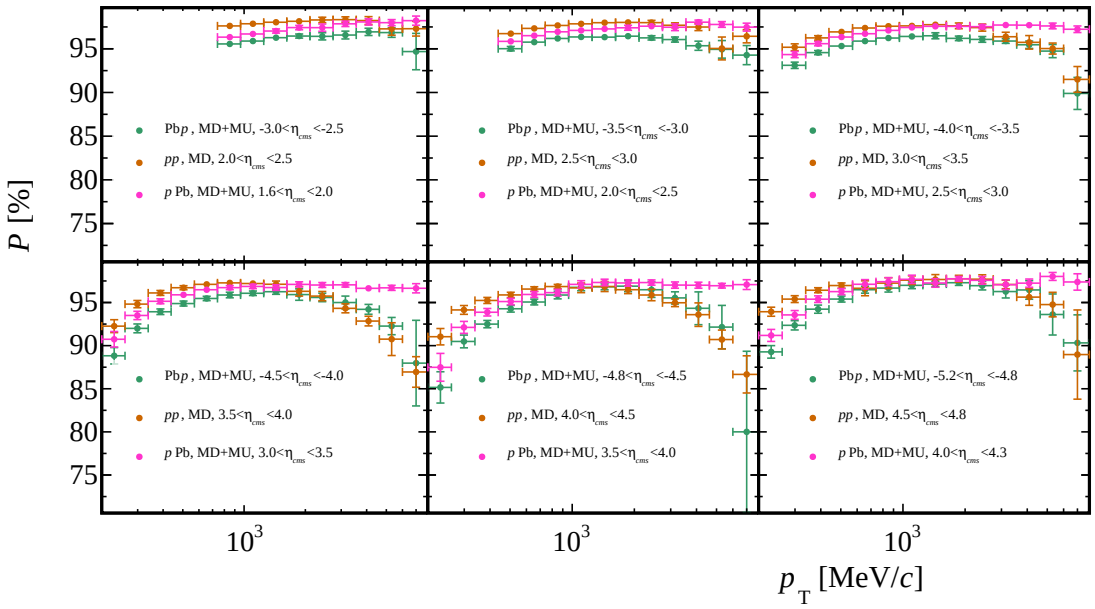


Figure 6.41: Final purity or fraction of no-background candidates for Pbp, pPb and pp obtained with Eq. 6.22. Error bars include systematic uncertainties.

6.5 Truth matching efficiency

The truth matching algorithm is in charge of establishing a correspondence between a charged particle, which is a generator level object, with a track, which is reconstruction level object. The truth matching efficiency is defined in Eq. 6.5. An efficiency below 100% means that some particles are reconstructed but the matching algorithm fails to link the MC particle with the reconstructed track. When this is the case, a generated prompt charged particle is wrongly considered as not reconstructed and a legitimate candidate track is classified as fake. To correct for this effect in the evaluation of the yield of charged tracks, N^{ch}, and in the fraction of fake tracks in simulation, $f^{\mathrm{sim}}_{\mathrm{fake}}$, a truth matching efficiency, $\varepsilon_{\mathrm{TM}}$, is included in Eqs. 6.1 and 6.13, respectively.

The truth matching efficiency complement is the fraction of unmatched signal candidates in simulation ,

$$f_{\mathrm{misfake}} = 1 - \varepsilon_{\mathrm{TM}}. \tag{6.23}$$

To obtain f_{misfake} a similar strategy as the one used to correct the selection efficiency in Sec. 6.3 is followed. Using a tag-and-probe approach with the $\phi(1020) \to K^{\pm}K^{\mp}$ decay, a calibration sample of $K^{\pm}$ is extracted. This sample is employed as proxy for signal charged particles. The candidate requirements are considered in the probe kaons and the

fraction of unmatched signal candidates is obtained,

$$f_{\text{misfake}} = \frac{N(K^{\pm}_{probe})|_{non-matched\&\&candidates}}{N(K^{\pm})|_{candidates}}. \tag{6.24}$$

This expression can be applied considering different kinematic bins in pseudorapidity and transverse momentum for the probe kaon in order to have a better description of ε_{TM}. The procedure assumes that the truth matching efficiency of a probed kaon of $(\eta^{K^{\mp}}, p_{\text{T}}^{K^{\mp}})$ is the same as for any other charged particle with the same (η, p_{T}) values. Due to the reduced number of miss-matched tracks, only four kinematic bins are considered.

In Figs. 6.42, 6.43 and 6.44, the $\phi(1020) \to K^{\pm}K^{\mp}$ signal extraction and the unmatched signal and background components are shown for proton-lead and *pp* configurations, respectively. The corresponding results of ε_{TM} are presented in Tab. 6.6. The table shows the result for each kinematic bin along with their statistical uncertainty propagated from the signal yields. Additional sources of uncertainty are neglected, as they would not have an impact in the measurement given the smallness of the correction.

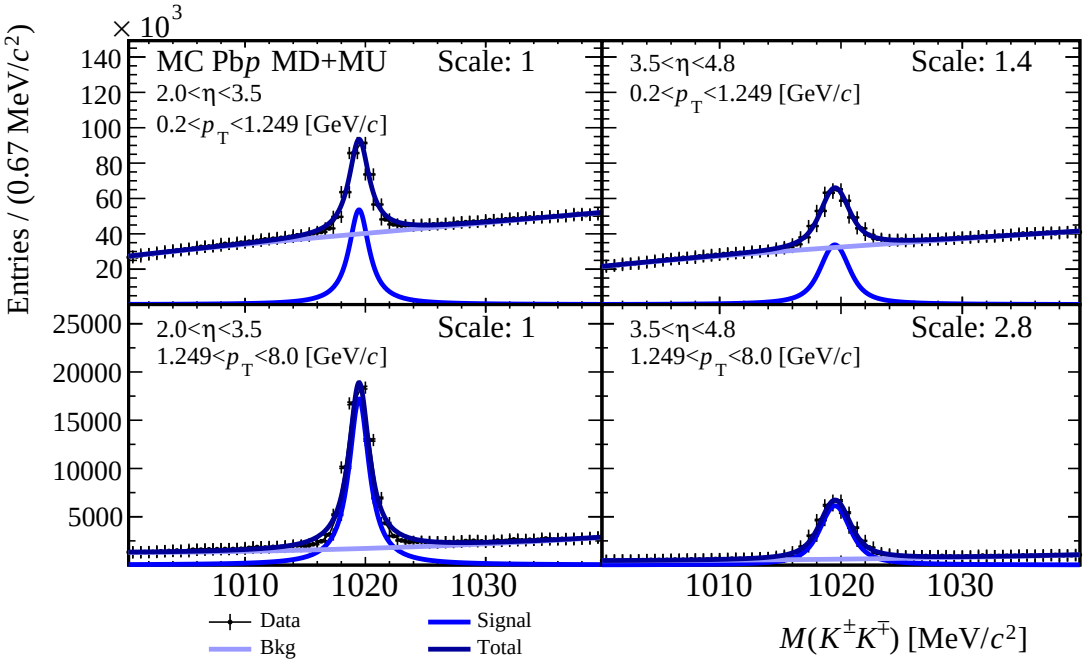
An important remark is that the truth-matching efficiency depends on the candidate selection. If no selection requirements are applied to the probe kaon slightly lower truth-matching efficiencies are found. These values are shown for reference in Tab. 6.7, although they are not used in the final result. This is a expected behaviour, since good quality tracks are more likely to meet the requirements of the truth-matching algorithm. The fraction of good quality tracks is significantly increased after applying the `GhostP` requirement. The same observation explains that the highest truth-matching efficiency occurs in the *pp*, then in the *p*Pb and finally in the Pb*p* configurations.

$p_{\text{T}}\,[\text{GeV}/c];\ \eta$	$\varepsilon_{TM,pp}$ (MCsim09d)	$\varepsilon_{TM,p\text{Pb}}$ (MCsim09e)	$\varepsilon_{TM,\text{Pb}p}$ (MCsim09e)
$0.2 < p_{\text{T}} < 1.249; 2 < \eta < 3.2$	0.9944 ± 0.0015	0.99029 ± 0.00056	0.98942 ± 0.00070
$0.2 < p_{\text{T}} < 1.249; 3.2 < \eta < 4.8$	0.9913 ± 0.0012	0.98873 ± 0.00059	0.98785 ± 0.00062
$1.249 < p_{\text{T}} < 8; 2 < \eta < 3.2$	0.9903 ± 0.0017	0.98708 ± 0.00014	0.98358 ± 0.00082
$1.249 < p_{\text{T}} < 8; 3.2 < \eta < 4.8$	0.9870 ± 0.0021	0.985514 ± 0.000082	0.98267 ± 0.00012

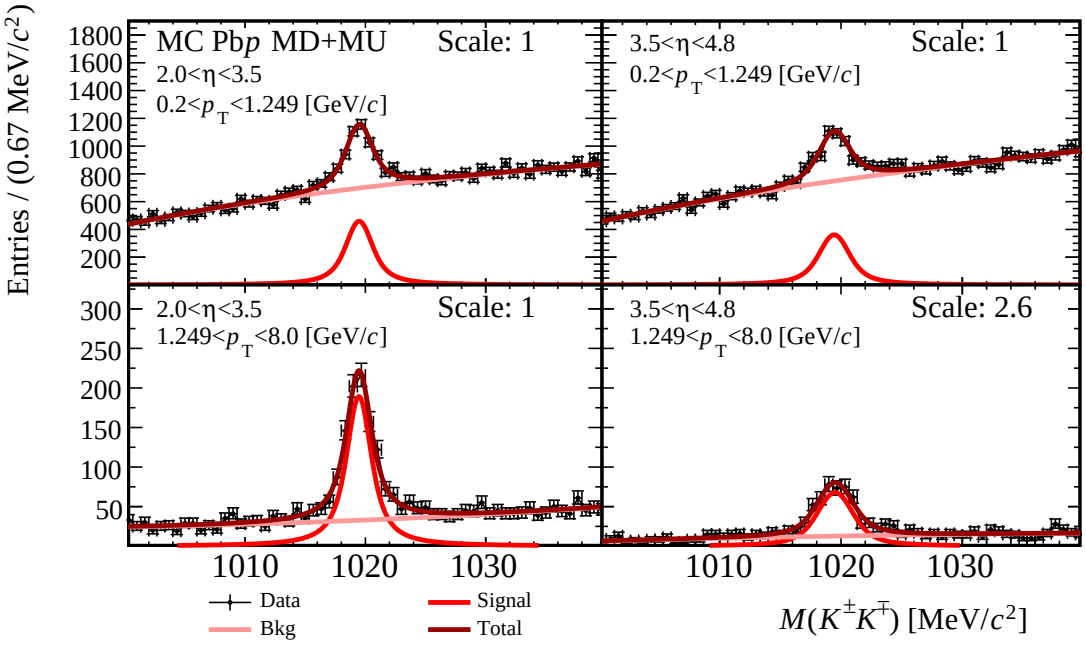
Table 6.6: Value of ε_{TM} for the Pb*p*, *p*Pb and *pp* simulation samples. The values are obtained with Eq. 6.24, imposing the selection requirements to the probed kaon.

$p_{\text{T}}\,[\text{GeV}/c];\ \eta$	$\varepsilon_{TM,pp}$ (MCsim09d)	$\varepsilon_{TM,p\text{Pb}}$ (MCsim09e)	$\varepsilon_{TM,\text{Pb}p}$ (MCsim09e)
$0.2 < p_{\text{T}} < 1.249; 2 < \eta < 3.2$	0.9886 ± 0.0023	0.9852 ± 0.0010	0.97640 ± 0.00095
$0.2 < p_{\text{T}} < 1.249; 3.2 < \eta < 4.8$	0.9830 ± 0.0022	0.9759 ± 0.0013	0.9738 ± 0.0048
$1.249 < p_{\text{T}} < 8; 2 < \eta < 3.2$	0.9818 ± 0.0028	0.9755 ± 0.0014	0.9754 ± 0.0017
$1.249 < p_{\text{T}} < 8; 3.2 < \eta < 4.8$	0.9804 ± 0.0030	0.9597 ± 0.0019	0.9618 ± 0.0020

Table 6.7: Value of ε_{TM} for the Pb*p*, *p*Pb and *pp* simulation samples. The values are obtained with Eq. 6.24, imposing the selection requirements to the probed kaon, without imposing the selection requirements to the probed kaon.

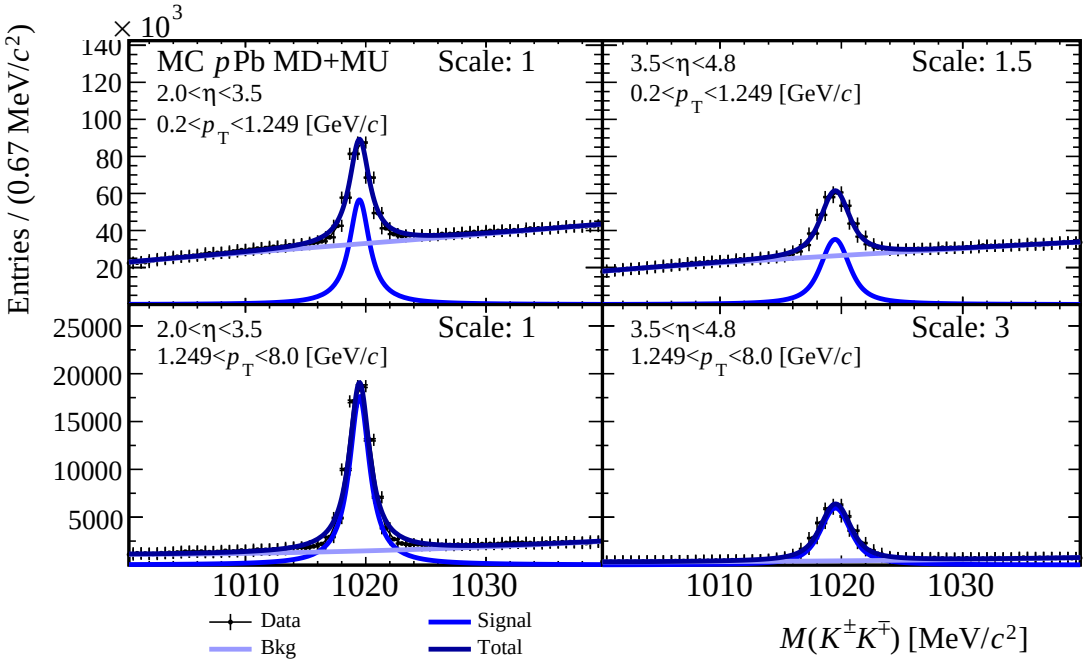


(a)

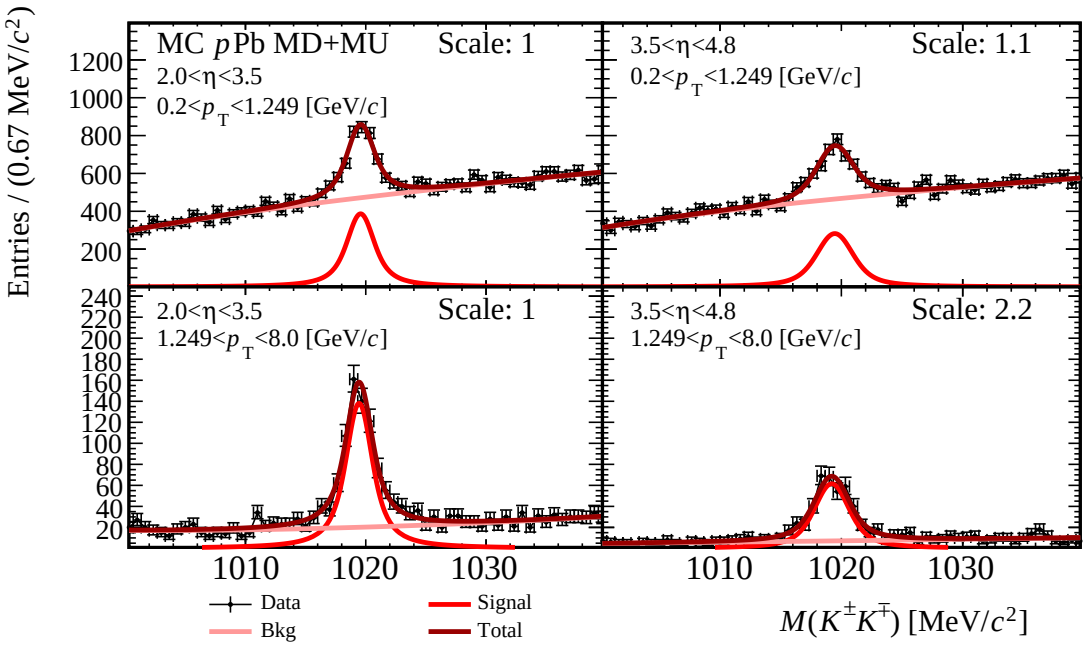


(b)

Figure 6.42: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ to compute f_{misfake} in the Pbp configuration. Four kinematic bins of $(\eta^{K^{\mp}}, p_{\text{T}}^{K^{\mp}})$ are considered. Top plots (Fig. 6.42a) shows the signal extraction considering all probe kaons, while bottom plots (Fig. 6.42a) show the fake track contribution (red). Plots are scaled by the factor indicated in the top right corner to ease visualisation.

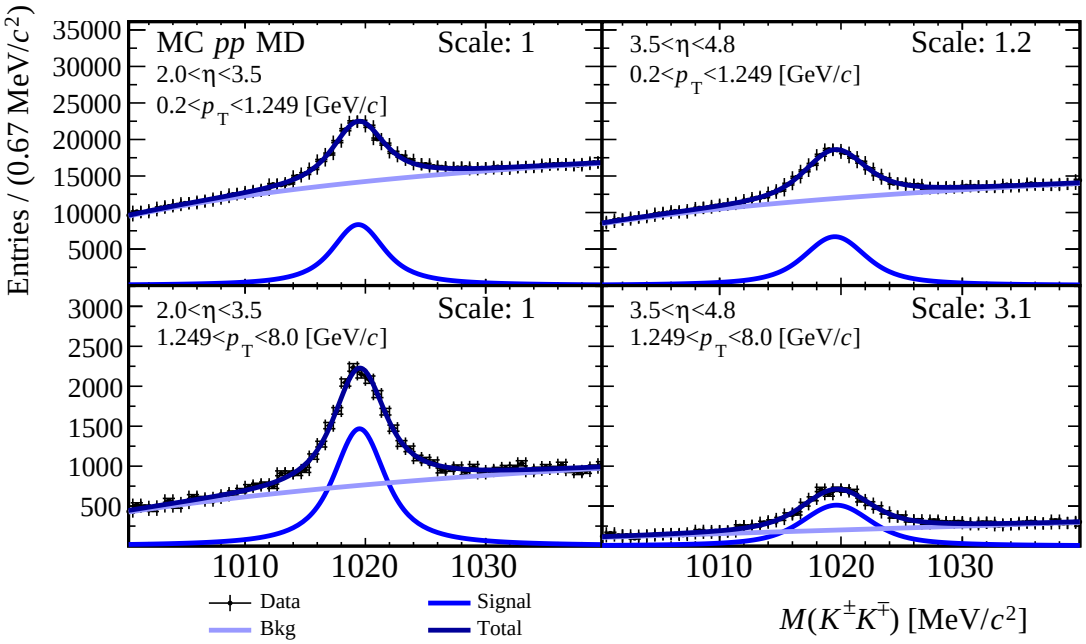


(a)

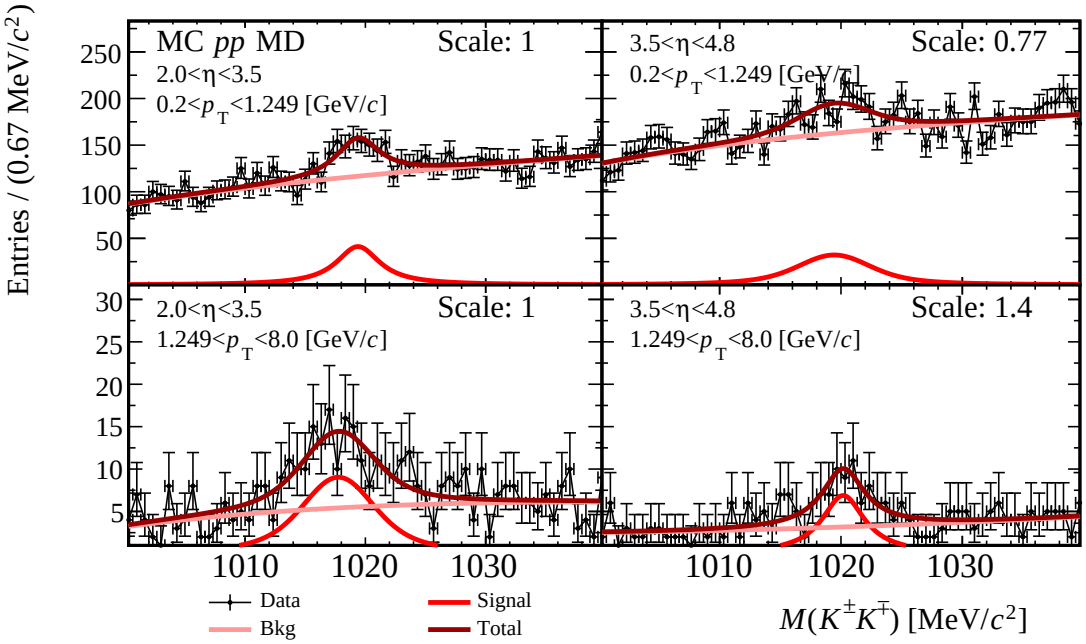


(b)

Figure 6.43: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ to compute f_{misfake} in the pPb configuration. Four kinematic bins of $(\eta^{K^{\mp}}, p_{\text{T}}^{K^{\mp}})$ are considered. Top plot (Fig. 6.43a) shows the signal extraction considering all probe kaons, while bottom plots (Fig. 6.43b) show the fake track contribution (red). Plots are scaled by the factor indicated in the top right corner to ease visualisation.



(a)



(b)

Figure 6.44: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ to compute f_{misfake} in the pp configuration. Four kinematic bins of $(\eta^{K^{\mp}}, p_{\text{T}}^{K^{\mp}})$ are considered. Top plot (Fig. 6.44a) shows the signal extraction considering all probe kaons, while bottom plots (Fig. 6.44b) show the fake track contribution (red). Plots are scaled by the factor indicated in the top right corner to ease visualisation.

6.6 Acceptance correction in proton-lead collisions

To compute $R^{\rm ch}_{p\rm Pb}$ the production cross-sections of pPb, Pbp and pp collisions must be measured in the centre-of-mass reference frame. In the LHC proton-lead collisions the energy of the lead beam ($E_{\rm Pb} = 1.58\,\text{TeV}$ per nucleon) is different to the energy of the proton beam ($E_p = 4\,\text{TeV}$). This means a nucleon-nucleon centre-of-mass energy of $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$. Due to the beam momentum asymmetry the centre-of-mass frame moves in the direction of the proton beam in the laboratory. For this reason, to obtain the centre-of-mass rapidity ($\eta_{\rm cms}$), it is necessary to introduce a boost of $y_{\rm boost} = -0.465$ in the direction of the proton beam to the measured laboratory pseudorapidity (η) in the proton-lead samples. For massless particles, the pseudorapidity is equal to the rapidity, and the relation

$$\eta_{\rm cms} = \eta + y_{\rm boost}. \tag{6.25}$$

For massive particles, the equivalence between rapidity and pseudorapidity is not linear and a more complex relationship holds,

$$\sinh\eta_{\rm cms} = \sinh(\eta - y_{\rm boost}) - \left(\sqrt{\frac{m^2}{p_{\rm T}^2} + \cosh^2\eta} - \cosh\eta\right)\sinh\eta_{\rm cms}. \tag{6.26}$$

Here, m is the mass of the particle and $p_{\rm T}$ is its transverse momentum. For most of the kinematic range under study, $p_{\rm T} >> m$, and thus the second term in the right side of Eq. 6.26 can be neglected and Eq. 6.25 is a good approximation. The $p_{\rm T}$ is unchanged by the boost as it only affects the longitudinal direction.

The LHCb acceptance for charged particles is $2 < \eta < 4.8$ and $p > 2\,\text{GeV}/c$. Considering Eq. 6.25, the acceptance in the centre-of-mass system translates to $1.5 < \eta_{\rm cms} < 4.3$ for pPb and $-5.3 < \eta_{\rm cms} < -2.5$ for Pbp. In Tab. 6.1 the equivalency between bins in η_{cms} and η was shown. However, there is an effect to the spectra caused by particles moving in and out of the acceptance according to their exact boost given by Eq. 6.26. The shift depends on the value of their $m/p_{\rm T}$ ratio. For a proton ($m = 938\,\text{MeV}/c^2$) of $p_{\rm T} = 200\,\text{MeV}/c$ at $\eta = 2$, the difference between $\eta_{\rm cms}$ and $\eta + y_{boost}$ is of about 6.4%, while for kaons ($m = 494\,\text{MeV}/c^2$) the difference is 2.3% and for a pion ($m = 140\,\text{MeV}/c^2$) is 0.2%. As η and $p_{\rm T}$ increase, the effect becomes smaller.

To study this effect, the generated distribution of charged particles $N^{gen}(\eta, p_{\rm T})$ in the EPOS simulation has been used. Only particles generated inside the LHCb acceptance are considered. From this distribution, the boosted spectra $N^{boost}(\eta_{\rm cms}, p_{\rm T})$ can be composed by performing a Lorentz boost to each generated particle. The *acceptance correction factor* can be defined

$$C_{\rm boost} = \frac{N^{boost}(\eta_{\rm cms}, p_{\rm T})}{N^{gen}_{shifted}(\eta, p_{\rm T})}, \tag{6.27}$$

where $N^{gen}_{shifted}(\eta, p_{\rm T})$ is the generated yield with the pseudorapidity of the particles has been shifted by $y_{\rm boost} = -0.465$. This correction quantifies the impact of using the

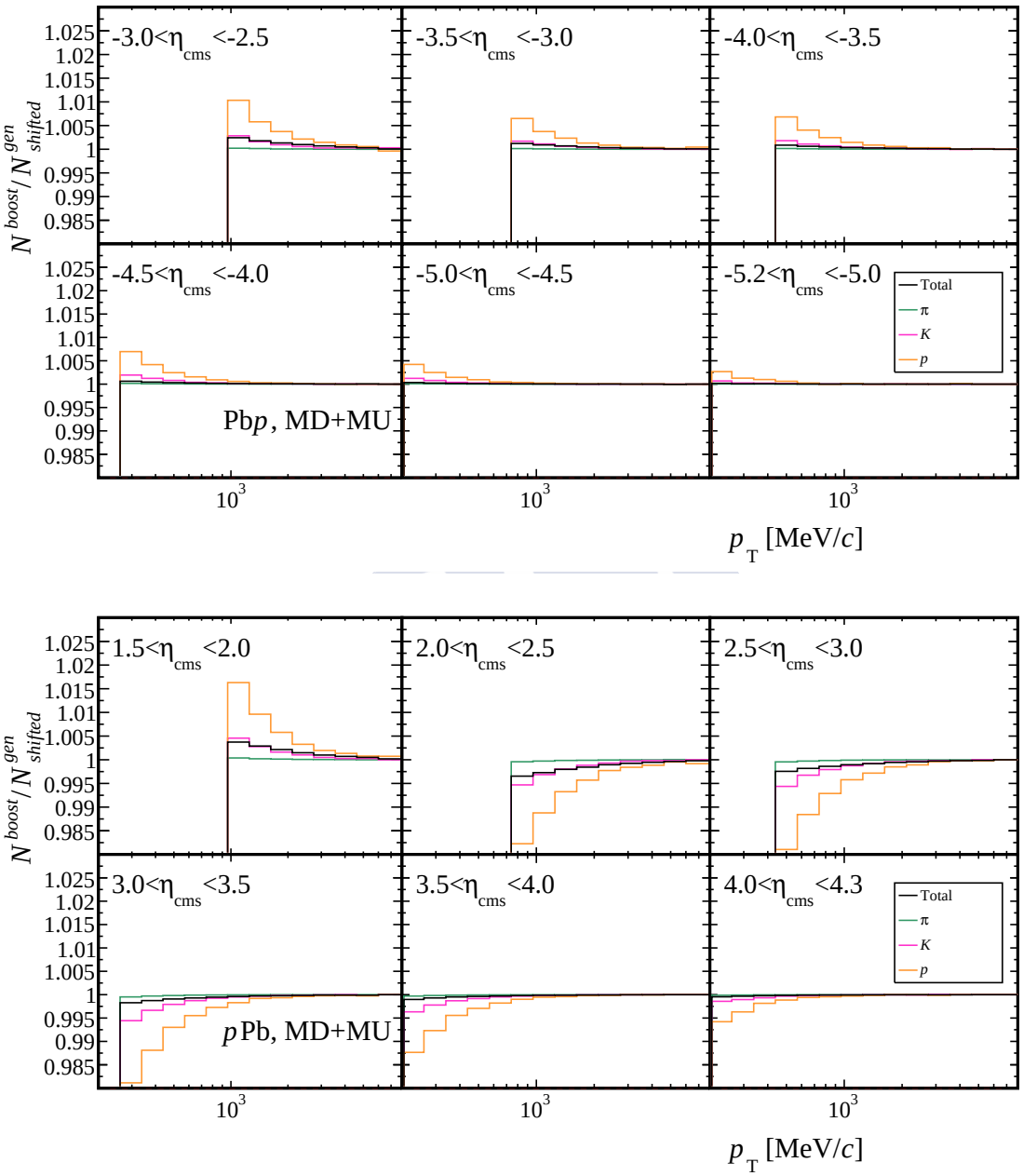


Figure 6.45: Acceptance correction in for each kinematic bin in the Pbp (top) and pPb (bottom) configurations. The correction is shown considering the full spectra, only pions, only kaons and only protons.

approximate Eq. 6.25 or the exact Eq. 6.26. In Fig. 6.45, this value is shown separately for protons, pions, kaons and for the full charged particle spectra. The correction is

around relevant for protons at low p_{T}, but not for pions, which constitute most of the spectra. For this reason, the correction for the combined spectra is smaller than 0.5% for every bin.

6.7 Bin migration

Bin migration consist in the assignment of reconstructed tracks to (η, p_{T}) bins that do not correspond with the real pseudorapidity and transverse momentum of the particle. The bin migration effect becomes more important when detector resolution is worse and when binning size is small.

The simulation samples can be used to study the impact of the bin migration effect. In the following, the measured value of pseudorapidity and transverse momentum will be referred to as $(\eta^{reco}, p_{\mathrm{T}}^{reco})$. The true value, denoted as $(\eta^{true}, p_{\mathrm{T}}^{true})$, is not accessible in the real data sample, but it can be studied in the simulation sample with generator level information. To study the resolution, the variables

$$\Delta p_{\mathrm{T}}/p_{\mathrm{T}} = \frac{p_{\mathrm{T}}^{true} - p_{\mathrm{T}}^{reco}}{p_{\mathrm{T}}^{true}}; \quad \Delta\eta/\eta = \frac{\eta^{true} - \eta^{reco}}{\eta^{true}} \tag{6.28}$$

are defined. The distributions $\Delta p_{\mathrm{T}}/p_{\mathrm{T}}$ and $\Delta\eta/\eta$ measure the relative difference between the true and the reconstructed value of p_{T} and η, respectively. The RMS value of these distributions is a good proxy for the detector resolution for charged particles. In Fig. 6.46, this RMS values are shown for the η and p_{T} bins of the analysis. The worst resolution occurs at very low p_{T}, and then increases rapidly reaching its maximum around $2\,\mathrm{GeV}/c$ with an RMS $\approx 0.8\%$ at around $2\,\mathrm{GeV}/c$ and then increases for larger p_{T}. Resolution is better in η, worsening for high η.

The p_{T} bin migration is studied by representing the values of p_{T}^{reco} versus p_{T}^{true} for every (η, p_{T}) bin. This is shown in Fig. 6.47 for pPb and Pbp, and in Fig. 6.48 for pp. The effect seems similar in every configuration, and the loss of entries in the diagonal does not exceed 5% in most of the bins, which indicates a very small effect from bin migration.

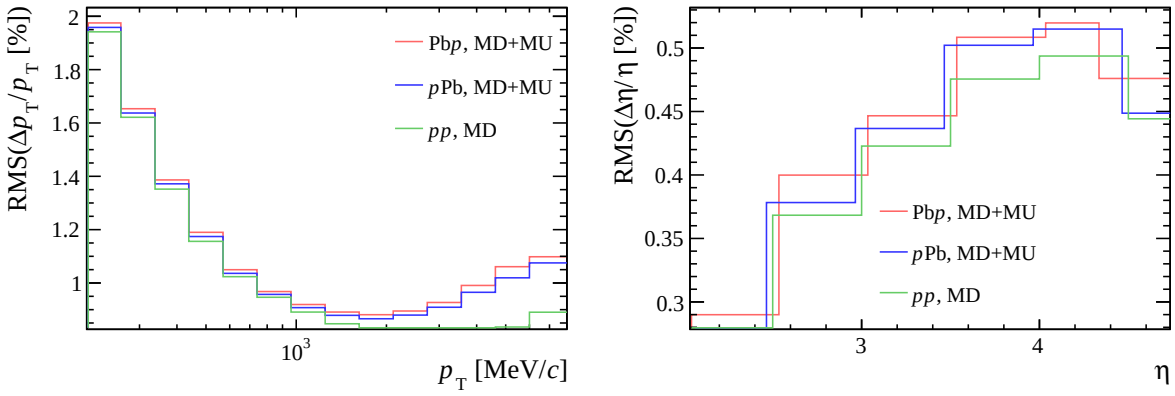
To quantify the magnitude of the effect, an unfolding procedure can be used. The generated spectra N^{true} and the reconstructed spectra N^{reco} are connected as,

$$N^{true}(\eta^{true}, p_{\mathrm{T}}^{true}) \xrightarrow{\mathbf{R}} N^{reco}(\eta^{reco}, p_{\mathrm{T}}^{reco}), \tag{6.29}$$

were $\mathbf{R}$ is the response matrix, which encapsulates the detector response. The idea of unfolding is to compute the matrix that performs the reversed process, which is called the unfolding matrix $\mathbf{U}$. This matrix allows to recover the N^{gen} distribution from N^{reco} as

$$N^{reco}(\eta^{reco}, p_{\mathrm{T}}^{reco}) \xrightarrow{\mathbf{U}} N^{true}(\eta^{true}, p_{\mathrm{T}}^{true}). \tag{6.30}$$

The first step is to construct the spectra $N^{reco}(\eta^{reco}, p_{\mathrm{T}}^{reco})$ and $N^{true}(\eta^{true}, p_{\mathrm{T}}^{true})$ with reconstructed and selected prompt charged particles inside the LHCb acceptance, using



(a) (b)

Figure 6.46: Resolution of LHCb in p_{T} and η for matched candidates. The showed bins correspond to the (η, p_{T}) bins covered in the analysis.

the reconstructed variables in the first case and the true variables in the second. These two distributions can be used to extract the response matrix.

A bin migration correction can be obtained at first approximation as the ratio between N^{true} and N^{reco}, where the first is in the variables $(\eta^{true}, p_{\mathrm{T}}^{true})$ and the second in $(\eta^{reco}, p_{\mathrm{T}}^{reco})$. This is equivalent to a bin-by-bin unfolding. This ratio is shown in Fig. 6.49 for the three data samples. The observed effect is reduced up to 0.5% for all the (η, p_{T}) bins in every beam configuration.

A more advance approach is to perform a proper unfolding of the distribution of reconstructed candidates. The idea is to remove from the distribution of candidates obtained with real data the influence of the bin migration effect. This is the distribution $N_{\mathrm{candidates}}$ that appears in Eq. 6.1 as input to obtain the yield. As the distributions N^{reco} and N^{true} that were used to extract the response matrix do not contain the contribution of background tracks, the background free distribution will be used, defined as,

$$N^{nobkg} = P^{sim} N_{\mathrm{candidates}}, \tag{6.31}$$

where P^{sim} was defined in Eq. 6.12 as the purity of the distribution of candidates in simulation. The previous equation is valid for the real data sample substituting the purity to the one in real data P as defined in Eq. 6.22.

There are several options to perform the unfolding procedure. Some of these methods are implemented in the RooUnfold framework [173] [174], integrated in ROOT. The latest version is available in a public repository [175]. In this analysis the Bayesian unfolding technique [176], also called *expectation-maximisation iteration with early stopping*, is used. This method has as advantage that it can deal with two-dimensional problems such as the one under study. The regularisation parameter, which is the number of iterations of the procedure, is set to 20.

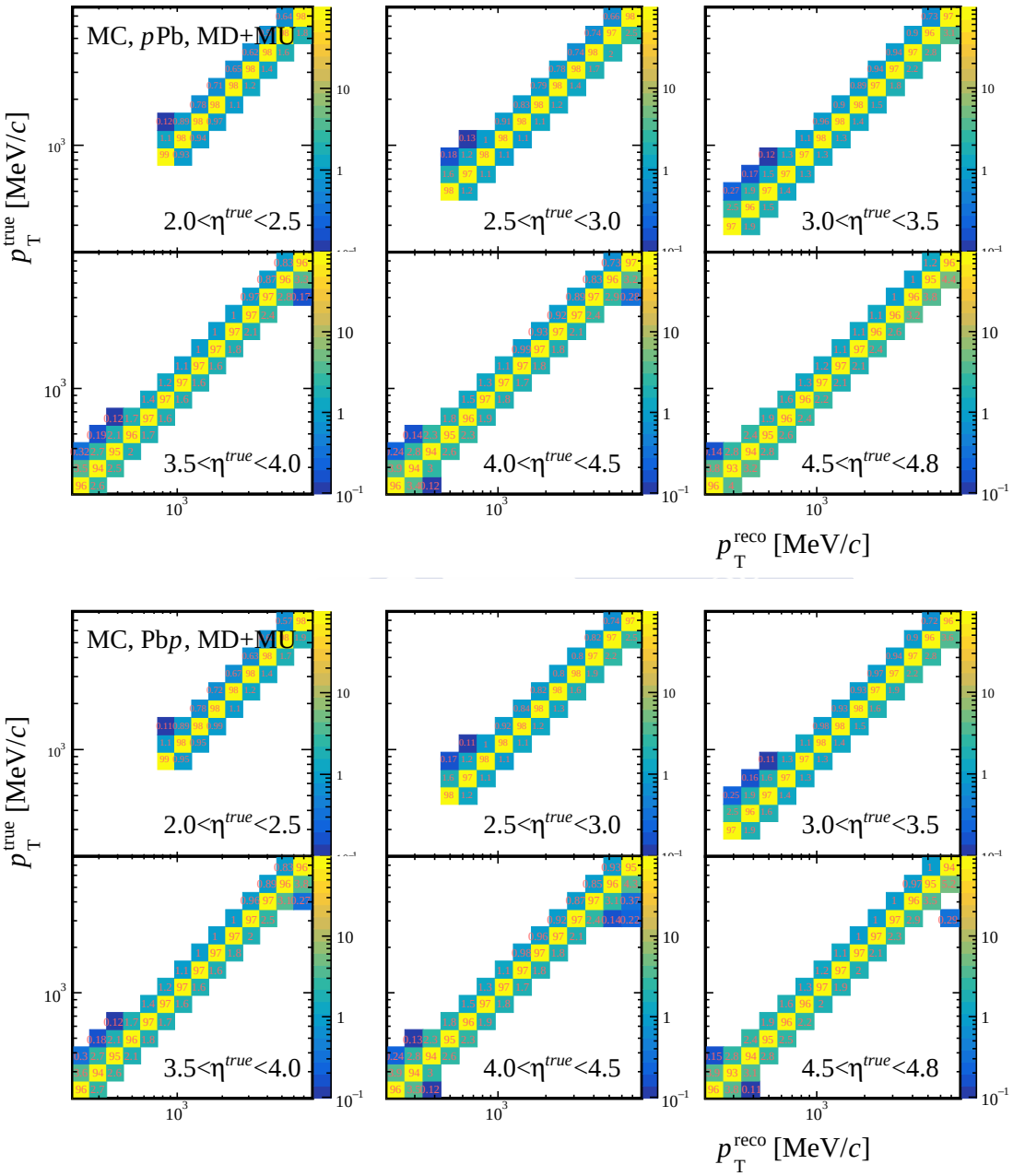


Figure 6.47: Effect of bin migration in p_{T} for Pbp (top) and pPb (bottom). Each bin show the fraction of particles in % with a given p_{T}^{reco} that are reconstructed with the value of p_{T}^{true} of the vertical axis. Only bins > 0.01 are shown.

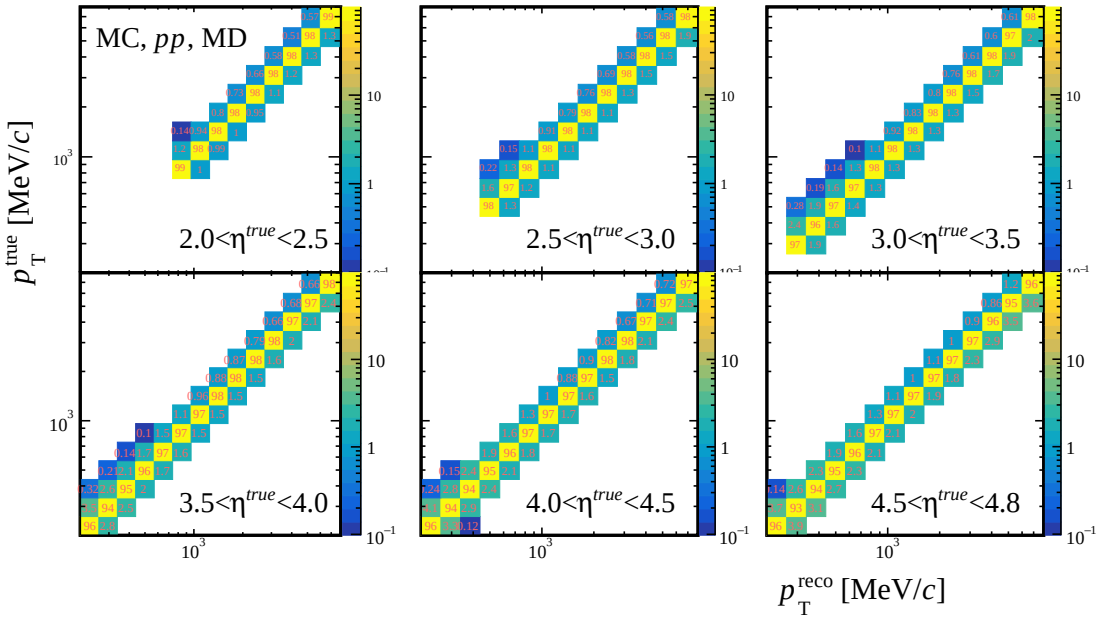


Figure 6.48: Effect of bin migration in p_{T} for pp configuration. Each bin show the fraction of particles in % with a given p_{T}^{reco} that are reconstructed with the value of p_{T}^{true} of the vertical axis. Only bins > 0.01 are shown.

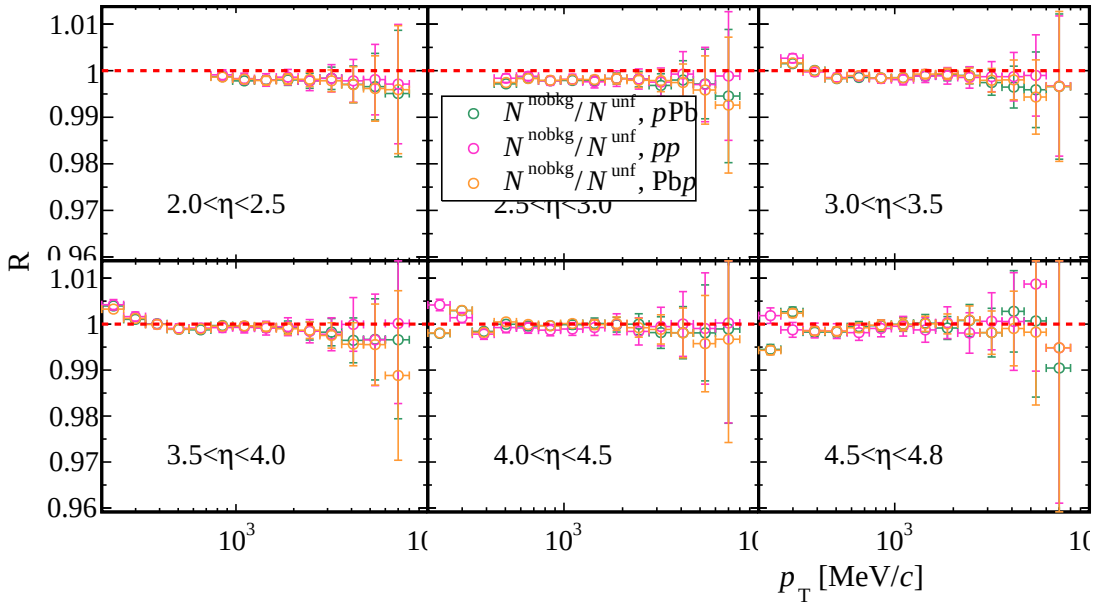


Figure 6.49: Ratio between the background-free reconstructed spectra and the generated spectra for Pbp, pPb and pp data. The deviation from unity is due to the bin migration effect.

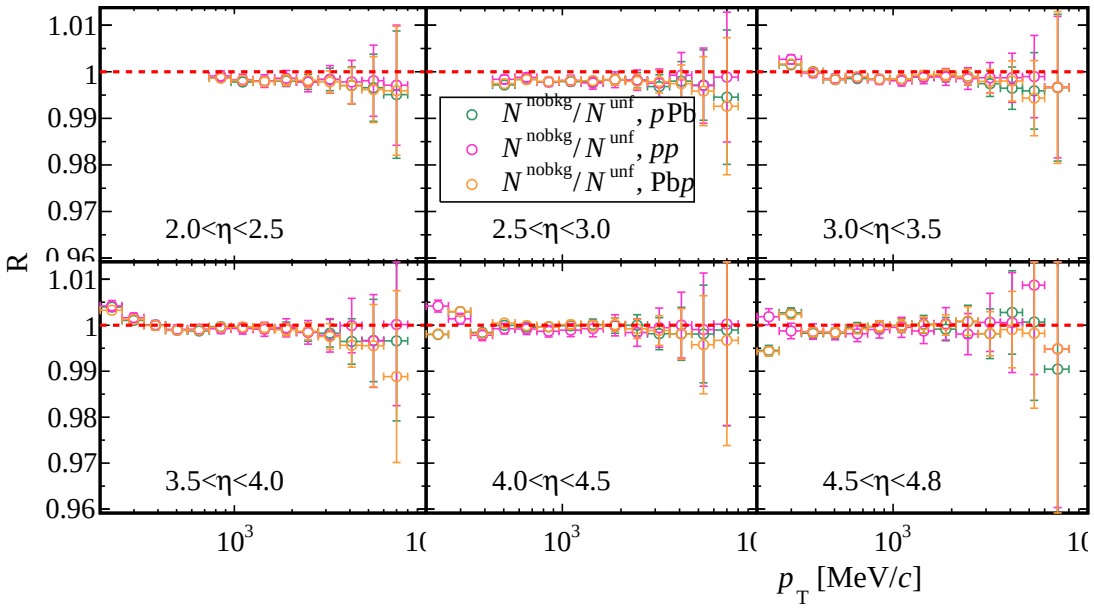


Figure 6.50: Ratio between the background free reconstructed spectra and the unfolded spectra for Pbp, pPb and pp data. The deviation from unity is due to the bin migration effect.

The unfolding procedure is validated with a closure test. This can be done by unfolding the N^{reco} distribution obtained with simulated data and check its compatibility with N^{true}. The ratio between these two distributions is known as the bias of the procedure, and should be unity. The result of this test shows no significant bias in the unfolding.

After validating the unfolding, we can finally check its effect in the background free distribution of candidates. This is shown in Fig. 6.50. The result is equivalent to the obtained with the bin-by-bin method presented in Fig. 6.49: no important effect is observed for any of the samples, apart from deviations lower than 0.5% in the border bins. Since these deviations are corrected by the bin-by-bin method, the effect is already corrected by $\varepsilon_{\text{reco}}$ and no unfolding correction is necessary.

6.8 Total correction

Taking Eq. 6.1, $\varepsilon_{\text{total}}$ can be defined as:

$$\varepsilon_{\text{total}} = \frac{\varepsilon_{\text{reco}}\varepsilon_{\text{sel}}(1/\varepsilon_{\text{TM}})}{PC_{\text{boost}}}. \tag{6.32}$$

After combining all the factors, results for $\varepsilon_{\text{total}}$ are shown in Fig. 6.51.

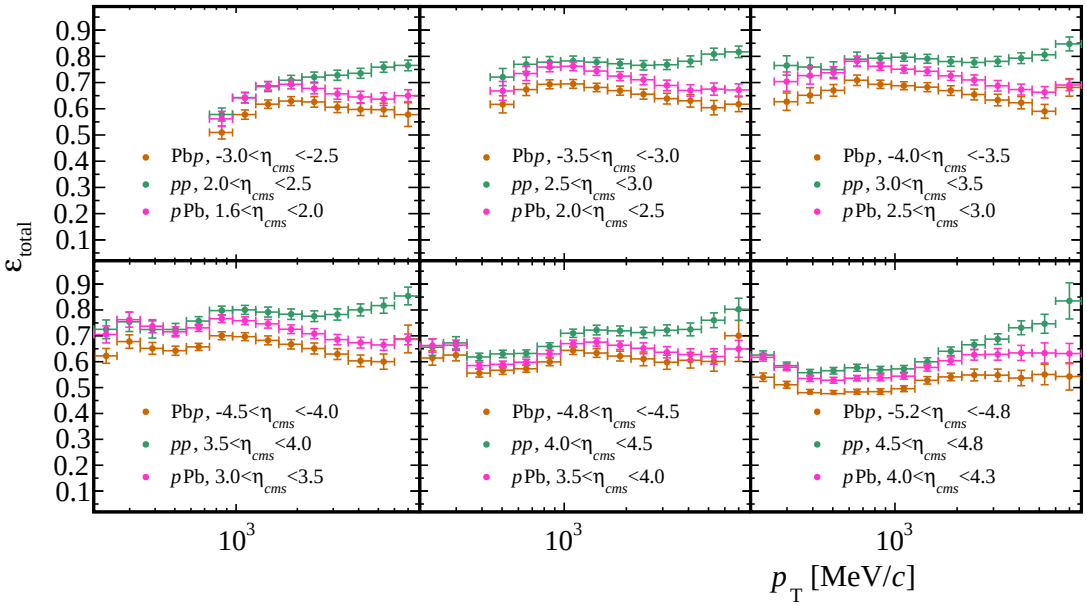


Figure 6.51: $\varepsilon_{\text{total}}$ in Pbp, pPb and pp dataset.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7

Systematic uncertainties

In this chapter the evaluation of the systematic uncertainties that affect the measurement of the observables defined in Sec. 2.5 is addressed. There are several sources of systematic uncertainty that affect Eq. 2.12. In general they are different for Pbp, pPb and pp and they depend on the (η, p_{T}) bin. From Sec. 7.1 to Sec. 7.5, the different contributions are computed and discussed. To conclude the chapter, a summary of the uncertainty of the measurement is presented in Sec. 7.6.

7.1 Reconstruction efficiency

There are three considered sources of systematic uncertainty affecting the reconstruction efficiency: the uncertainty in the tracking correction $C_{tracking}(p, \eta)$, the detector occupancy description and the particle composition of the simulation.

7.1.1 Tracking correction

The $C_{tracking}(p, \eta)$ correction factors computed to account for the data/simulation discrepancy in Sec. 6.2 are at the origin of some systematic uncertainties. Each (p, η) bin has a systematic uncertainty from calibration, as indicated in Figs. 6.2 and 6.5. A global systematic uncertainty of 0.4% (0.8%) is estimated for the proton-lead (proton-proton) data calibrated with the 2012 (2015) sample, see Secs. 6.2.1 and 6.2.2 for details. This applies to tracks within the calibrated $5 < p < 200\,\mathrm{GeV}/c$ and $2 < \eta < 4.8$ region. Low ($2 < p < 5\,\mathrm{GeV}/c$) and very high ($p > 200\,\mathrm{GeV}/c$) momentum non-calibrated

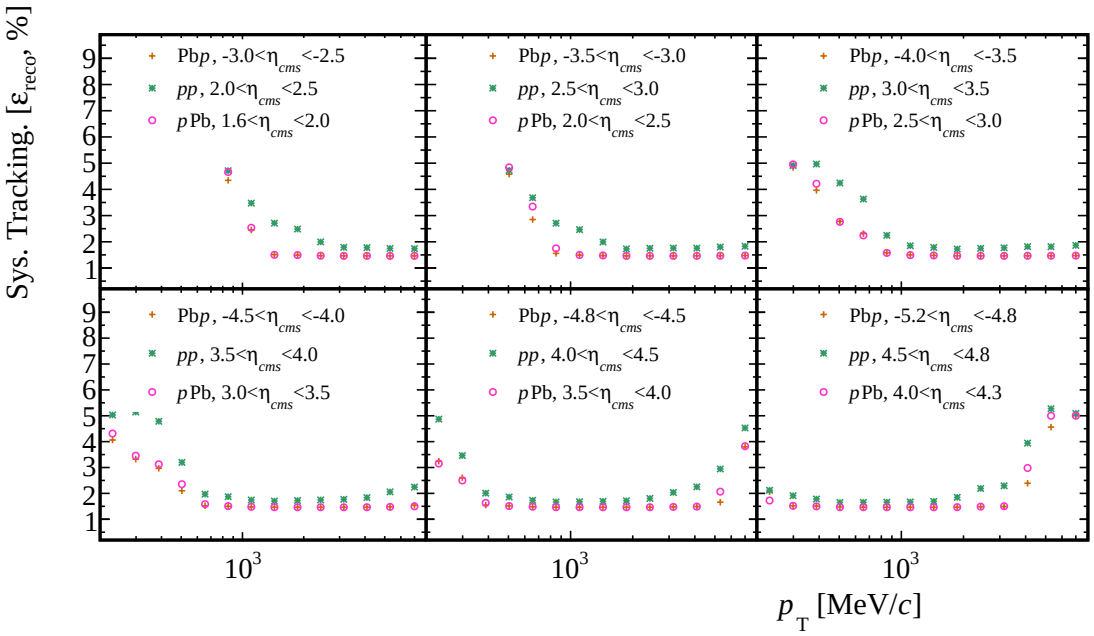


Figure 7.1: Relative systematic uncertainty in $\varepsilon_{\text{reco}}$ from the tracking correction $C_{tracking}$ for Pbp, pPb and pp data.

tracks are conservatively assigned a 5% systematic uncertainty. Moreover, since the calibration procedure employs muon tracks, an additional of 1.4% systematic uncertainty is assigned originated in the description of hadronic interactions with the detector material in simulation (see section 9 of [132] for more details).

To propagate this uncertainty in the calculation of $\varepsilon_{\text{reco}}$, given by Eqs. 6.6 and 6.7, the efficiency is computed varying $C_{tracking}(p,\eta)$ by $\pm\sigma$. The systematic uncertainty on $\varepsilon_{\text{reco}}$ is obtained as half the difference between these two values. To cross-check the result, the uncertainty have been obtained with an alternative method where 500 toys of $C_{tracking}(p,\eta)$ are generated with a gaussian distribution with the value of $C_{tracking}(p,\eta)$ as mean and the uncertainty of the bin as width. The reconstruction efficiency is calculated for every toy, and then, the $\varepsilon_{\text{reco}}$ distribution is fitted to a gaussian function for every $(\eta_{\text{cms}}, p_{\text{T}})$ bin. The final value of $\varepsilon_{\text{reco}}$ is given by the mean of the gaussian while the uncertainty is the width. Both methods give compatible results. The results for the relative systematic uncertainty in each $(\eta_{\text{cms}}, p_{\text{T}})$ interval are shown in Fig. 7.1.

7.1.2 Detector occupancy

Another systematic uncertainty originates in the weighting procedure which was used to improve the detector occupancy description of simulation.

In Sec. 5.2.1 it was concluded that the weight sets 1, 2, 9 and 10, obtained with different variable combinations, are equivalent in terms of description of the studied variables

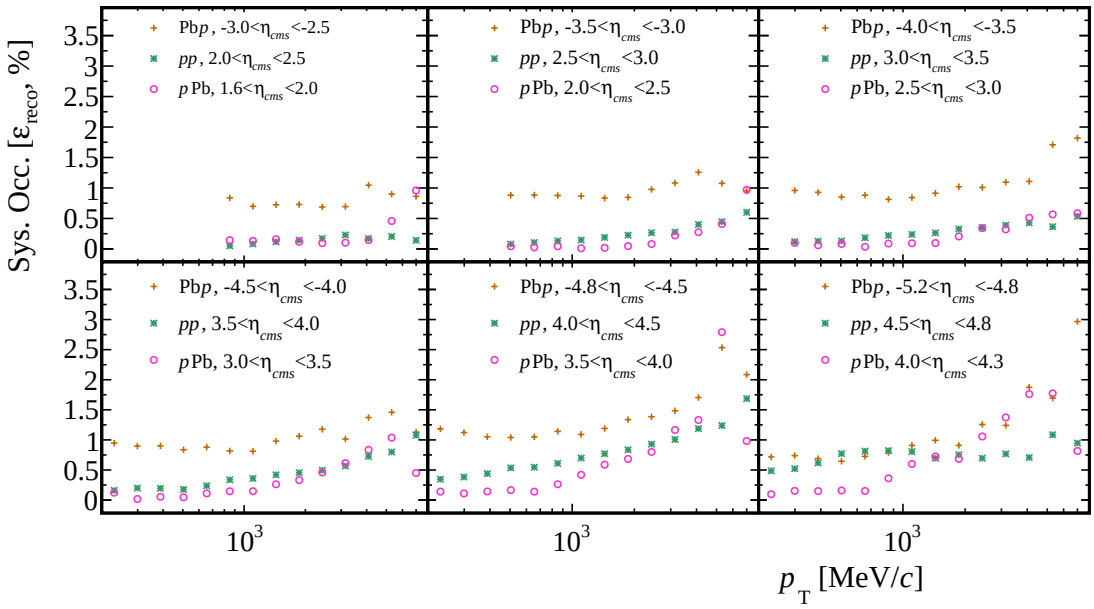


Figure 7.2: Relative systematic uncertainty in $\varepsilon_{\text{reco}}$ from detector occupancy description for Pbp, pPb and pp data.

for proton-lead. However, as it can be seen in Fig. B.7 for pPb and Fig. B.6 for Pbp, the reconstruction efficiency slightly depends on the chosen set.

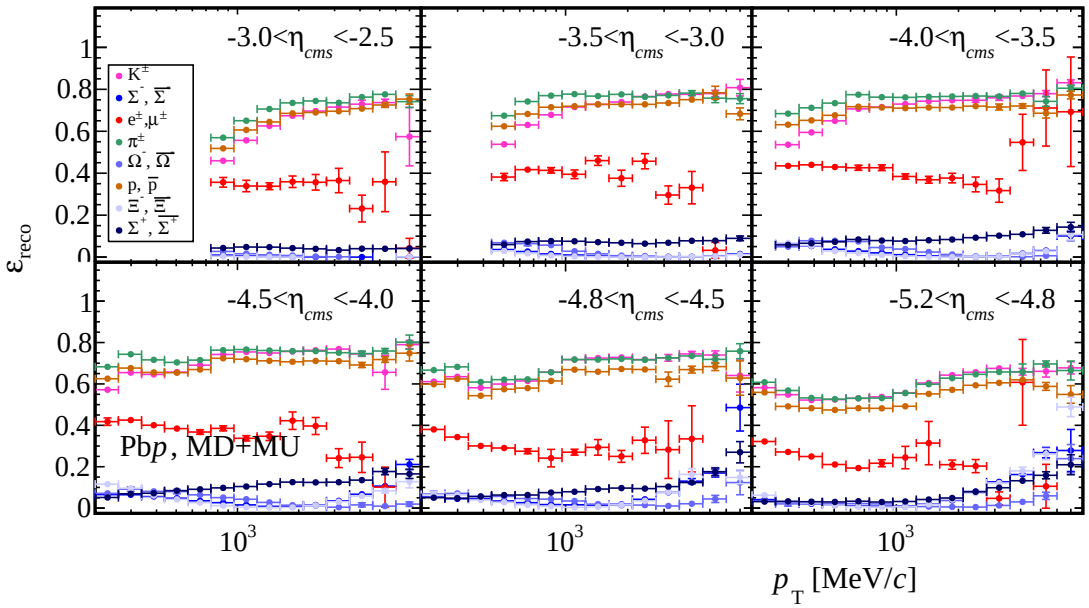
A systematic uncertainty is assigned for this reason. The weighted standard deviation of the values obtained with sets 1, 2, 9 and 10 is considered. The same procedure is followed for pp. The variation of $\varepsilon_{\text{reco}}$ with the different sets is shown in Fig. 6.7. In this case, sets 12, 13, 14 and 15 are used to compute the weighted standard deviation.

The relative systematic uncertainty from the detector occupancy obtained with this procedure is shown in Fig. 7.2. The uncertainty ranges from 0 to 3% and it is $\sim 1\%$ for most of the bins.

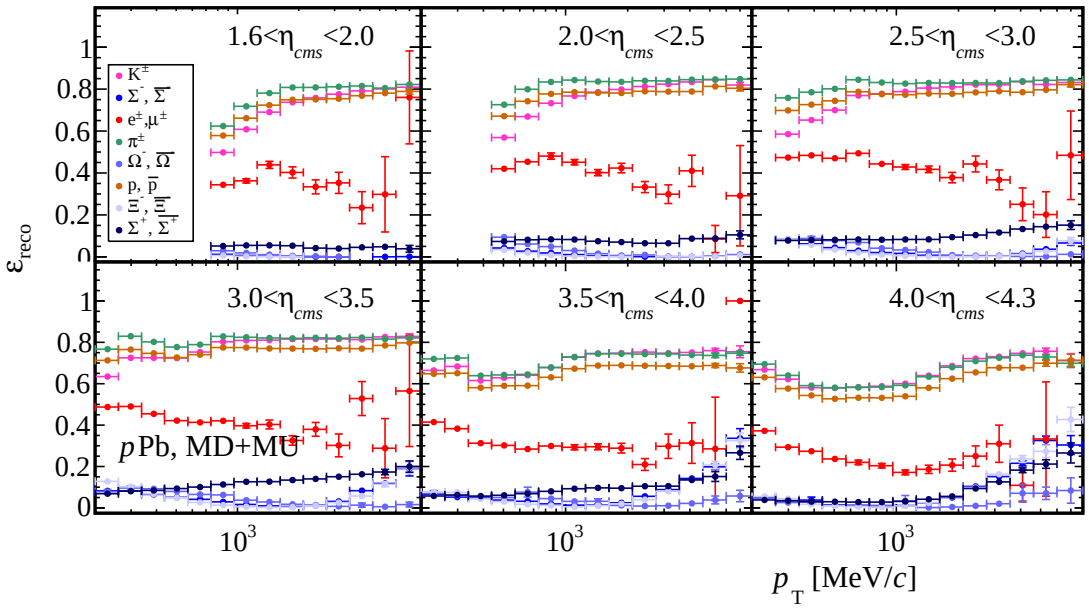
7.1.3 Particle composition

This systematic uncertainty accounts for differences of the prompt charged particle composition in the data and simulation samples used to compute the reconstruction efficiency. The study relies on the particle composition discussion of Sec. 5.4.3 and the hypotheses introduced therein.

Figs. 7.3 and 6.8 show the reconstruction efficiency by particle specie, obtained with Eq. 6.2 restricted to MC identified particles. The efficiency depends on the particle type: pions, kaons and protons have similar results, although some differences can be appreciated. These differences are attributed to the different lifetime of the particles and the different way they interact with the detector material. Muon and electron effi-



(a)



(b)

Figure 7.3: Reconstruction efficiency by charged particle species in Pbp (top) and pPb (bottom) configurations. Error bars shown statistical uncertainty.

ciencies are integrated in the figure and the presented efficiency is dominated by the more abundant electrons. However, both species do not significantly impact the global reconstruction efficiency due to their relative low abundance. For hyperons, the reconstruction efficiency is very low because they mostly decay before reaching the last tracking station after the magnet, which is a long track reconstruction requirement. A slight increase of this efficiency is seen at high p, as a result of the higher boost that allows more hyperons to meet this requirement.

Particle composition in simulation has been validated with the available data in Sec. 5.4.3. There, it was concluded that the EPOS generator for proton-lead and pp provides a reasonable description of the h/π ratios within a 30% margin. Since the official pp simulation was generated with PYTHIA, $\varepsilon_{\text{reco}}$ was modified to introduce the particle fractions computed in the EPOS pp sample according to Eq. 6.8.

A systematic uncertainty can be obtained considering Eq. 6.8, where f_p is defined as:

$$f_p^{\text{sim}} = \frac{N_p^{\text{sim}}}{\sum_p N_p^{\text{sim}}}, \quad \text{with}\, p = \pi^-, K^-, p, e^-, \mu^-, \Xi^-, \Sigma^+, \Sigma^-, \Omega^-, cc. \tag{7.1}$$

The procedure follows three steps for each (η, p_{T}) bin of pPb, Pbp and pp:

1. First, $n = 1000$ instances of each N_p are generated following independent gaussian distributions. The parameters of the gaussians are fixed to $(\mu_p, \sigma_p) = (N_p^{\text{sim}}, \Delta N_p)$, where $\Delta N_p = 0.3 N_p^{\text{sim}}$ according to the 30% variation.

2. For each instance a f_p is computed and $\varepsilon_{\text{reco}}$ is calculated with Eq. 6.8.

3. The distribution of $\varepsilon_{\text{reco}}$ is fitted to a gaussian. The resulting μ should recover the original $\varepsilon_{\text{reco}}$ and the systematic uncertainty will be the σ from the fit. Some fit examples are shown in Fig. 7.4.

The systematic uncertainty from particle composition for each configuration is shown in Fig. 7.5. The systematic ranges from 0.5 to 4%, being largest at high p_{T}. This is because the abundance of K, p and hyperons with respect to π rises with p_{T} (see Sec. 5.4.3), and therefore the efficiency becomes more sensitive to variations of f_p.

7.1.4 Summary

An additional contribution to the uncertainty of $\varepsilon_{\text{reco}}$ comes from the limited statistics of the simulation. The relative contribution of this source is shown in the appendix (Fig. B.19). This uncertainty is added in quadrature with the already mentioned three contributions. Fig. 7.6 shows the total relative systematic uncertainty of $\varepsilon_{\text{reco}}$.

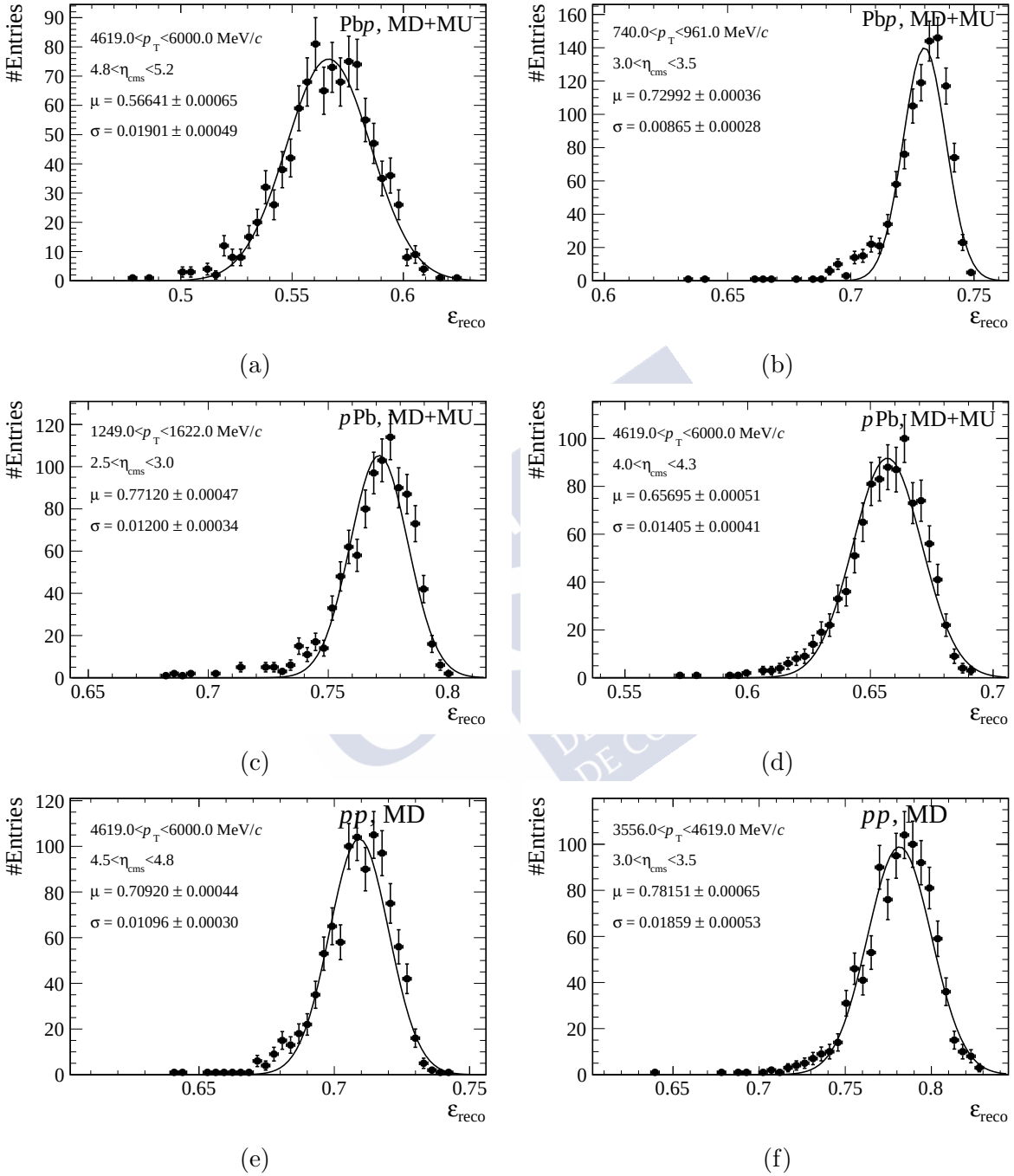


Figure 7.4: Gaussian fits for the study of the variation of $\varepsilon_{\text{reco}}$ with particle relative abundances for Pbp (top) pPb (middle) and pp (bottom) collisions for different kinematic bins.

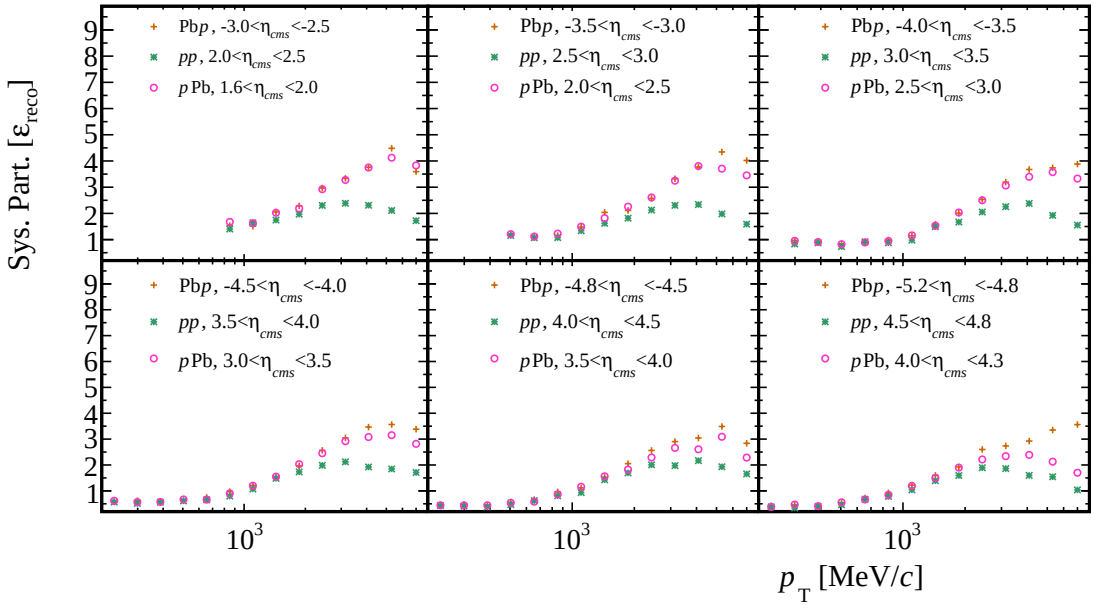


Figure 7.5: Relative systematic uncertainty in $\varepsilon_{\text{reco}}$ from particle composition for Pbp, pPb and pp data.

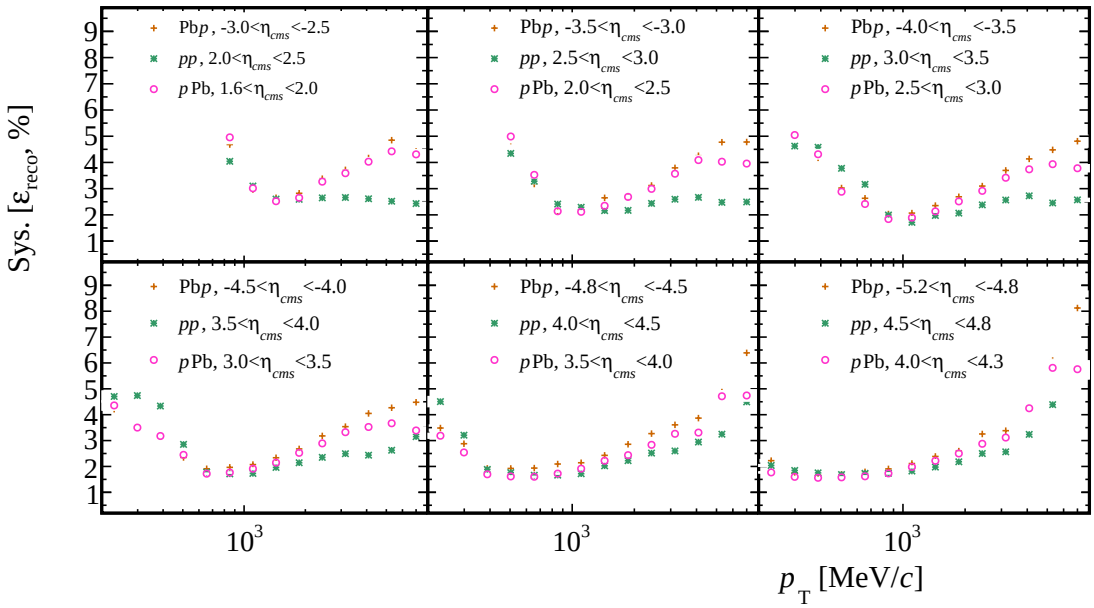


Figure 7.6: Relative systematic uncertainty in $\varepsilon_{\text{reco}}$ for Pbp, pPb and pp data.

7.2 Selection Efficiency

The systematic uncertainty in the selection efficiency mainly originates in the data-driven correction explained in Sec. 6.3. The first considered source of error is the $\phi(1020)$ yield uncertainty from the signal extraction. Yield results are detailed in Tab. 7.1 with the corresponding statistical uncertainty. This uncertainty is propagated to $\varepsilon_{sel}^{\text{data}}$ and $\varepsilon_{sel}^{\text{sim}}$ following Eq. 6.10, and then to the $C_{sel} = \varepsilon_{sel}^{\text{data}}/\varepsilon_{sel}^{\text{sim}}$ correction factor.

In addition, a systematic uncertainty due to the signal extraction procedure is assigned to C_{sel}. The uncertainty is computed by modifying the baseline background model in the signal extraction with a third order Chebychev polynomial. Every fit for the proposed kinematic bins of Pbp, pPb and pp for data and simulation is repeated, and C_{sel} is computed for the new values of $\varepsilon_{sel}^{\text{data}}$ and $\varepsilon_{sel}^{\text{sim}}$. The largest difference variation of C_{sel} with respect to the nominal value occurs for Pbp in the $(\eta, p_{\text{T}}) = ([3,4],[2,8]\,\text{GeV}/c)$ kinematic bin. The relevant fits are shown in Fig. 7.7. The difference corresponds to 0.8% of the value of C_{sel}, which is taken as systematic uncertainty.

The uncertainty from C_{sel} in the final selection efficiency is propagated with a similar technique as for $\varepsilon_{\text{reco}}$ and $C_{tracking}$. Eq. 6.9 is computed varying C_{sel} by $\pm\sigma$ and the systematic uncertainty is estimated as half the difference between the two results. This systematic is added in quadrature with the statistical uncertainty of ε_{sel}, which arises from the limited size of the simulation sample. This contribution is lower than 1% for all the bins and is presented in Fig. B.20.

The dependency of ε_{sel} with the particle composition of simulation is studied as for $\varepsilon_{\text{reco}}$. Fig. 7.8 shows the selection efficiency in the Pbp configuration for all particle species. This efficiency is obtained by applying Eq. 6.9 and requiring that particles have the corresponding ID. Clearly, ε_{sel} for π, K and p is the same all over the kinematic range, and therefore no systematic uncertainty needs to be assigned due to the unknown particle composition in the data sample. The behaviour in pPb and pp is very similar to Pbp. The efficiency is clearly different for hyperons, however this has no effect in ε_{sel}, because only reconstructed particles are accounted for in Eq. 6.9. From Figs. 7.3 and 6.8, the hyperons reconstruction efficiency is very small, which adds up to the small relative abundance of the hyperons. To conclude, the relative systematic uncertainty in ε_{sel} is shown in Fig. 7.9.

$p_{\mathrm{T}}\,[\,\mathrm{GeV}/c];\ \eta$	$N(\phi(1020))$, pPb			
	Data		Simulation	
	Before cuts	After cuts	Before cuts	After cuts
$0.2 < p_{\mathrm{T}} < 0.6; 2 < \eta < 3$	80900 ± 1400	74900 ± 1000	13440 ± 230	12910 ± 190
$0.2 < p_{\mathrm{T}} < 0.6; 3 < \eta < 4$	379000 ± 3200	347100 ± 2200	56450 ± 530	53900 ± 410
$0.2 < p_{\mathrm{T}} < 0.6; 4 < \eta < 4.8$	248700 ± 2800	223300 ± 1800	31970 ± 460	30320 ± 340
$0.6 < p_{\mathrm{T}} < 1; 2 < \eta < 3$	248900 ± 1800	231700 ± 1300	39980 ± 300	38680 ± 260
$0.6 < p_{\mathrm{T}} < 1; 3 < \eta < 4$	562600 ± 2600	524700 ± 1900	74010 ± 430	71300 ± 360
$0.6 < p_{\mathrm{T}} < 1; 4 < \eta < 4.8$	199100 ± 1600	185800 ± 1100	26510 ± 280	25670 ± 230
$1 < p_{\mathrm{T}} < 2; 2 < \eta < 3$	315900 ± 1500	296500 ± 1200	47830 ± 280	46300 ± 250
$1 < p_{\mathrm{T}} < 2; 3 < \eta < 4$	431200 ± 1700	406100 ± 1300	58270 ± 310	56290 ± 280
$1 < p_{\mathrm{T}} < 2; 4 < \eta < 4.8$	116720 ± 990	109660 ± 710	18420 ± 190	17780 ± 160
$2 < p_{\mathrm{T}} < 8; 2 < \eta < 3$	72490 ± 580	68440 ± 460	10030 ± 110	9720 ± 110
$2 < p_{\mathrm{T}} < 8; 3 < \eta < 4$	57180 ± 520	54010 ± 410	8620 ± 100	8316 ± 98
$2 < p_{\mathrm{T}} < 8; 4 < \eta < 4.8$	8390 ± 220	7920 ± 170	1633 ± 47	1569 ± 44
$p_{\mathrm{T}}\,[\,\mathrm{GeV}/c];\ \eta$	$N(\phi(1020))$, Pbp			
	Data		Simulation	
	Before cuts	After cuts	Before cuts	After cuts
$0.2 < p_{\mathrm{T}} < 0.6; 2 < \eta < 3$	67500 ± 1300	61770 ± 970	12300 ± 240	11800 ± 190
$0.2 < p_{\mathrm{T}} < 0.6; 3 < \eta < 4$	348800 ± 3600	314900 ± 2400	54550 ± 570	51930 ± 430
$0.2 < p_{\mathrm{T}} < 0.6; 4 < \eta < 4.8$	236400 ± 3300	209900 ± 2000	30520 ± 530	28780 ± 370
$0.6 < p_{\mathrm{T}} < 1; 2 < \eta < 3$	223500 ± 2000	205000 ± 1300	37940 ± 320	36560 ± 270
$0.6 < p_{\mathrm{T}} < 1; 3 < \eta < 4$	550100 ± 3100	501900 ± 2100	76620 ± 480	73180 ± 390
$0.6 < p_{\mathrm{T}} < 1; 4 < \eta < 4.8$	205500 ± 2000	187600 ± 1300	28240 ± 330	27050 ± 250
$1 < p_{\mathrm{T}} < 2; 2 < \eta < 3$	293500 ± 1700	269600 ± 1200	47190 ± 290	45440 ± 250
$1 < p_{\mathrm{T}} < 2; 3 < \eta < 4$	408500 ± 2000	377400 ± 1400	64120 ± 340	61550 ± 300
$1 < p_{\mathrm{T}} < 2; 4 < \eta < 4.8$	104100 ± 1100	95000 ± 740	20600 ± 210	19780 ± 180
$2 < p_{\mathrm{T}} < 8; 2 < \eta < 3$	58460 ± 620	54280 ± 420	9820 ± 110	9450 ± 110
$2 < p_{\mathrm{T}} < 8; 3 < \eta < 4$	42340 ± 530	39670 ± 360	8550 ± 110	8257 ± 99
$2 < p_{\mathrm{T}} < 8; 4 < \eta < 4.8$	4280 ± 170	3990 ± 120	1461 ± 47	1412 ± 42
$p_{\mathrm{T}}\,[\,\mathrm{GeV}/c];\ \eta$	$N(\phi(1020))$, pp			
	Data		Simulation	
	Before cuts	After cuts	Before cuts	After cuts
$0.2 < p_{\mathrm{T}} < 1; 2 < \eta < 3.5$	1139900 ± 4000	1072700 ± 3000	74600 ± 1100	71480 ± 810
$0.2 < p_{\mathrm{T}} < 1; 3.5 < \eta < 4.8$	1148100 ± 1700	1077200 ± 3300	74800 ± 1200	71580 ± 830
$1 < p_{\mathrm{T}} < 8; 2 < \eta < 3.5$	462100 ± 1600	438200 ± 1300	28000 ± 430	26840 ± 360
$1 < p_{\mathrm{T}} < 8; 3.5 < \eta < 4.8$	204000 ± 1100	195990 ± 870	12420 ± 300	12040 ± 230

Table 7.1: Yield results with the statistical uncertainty from the fit to extract the signal of $\phi \to K^{\pm}K^{\mp}$.

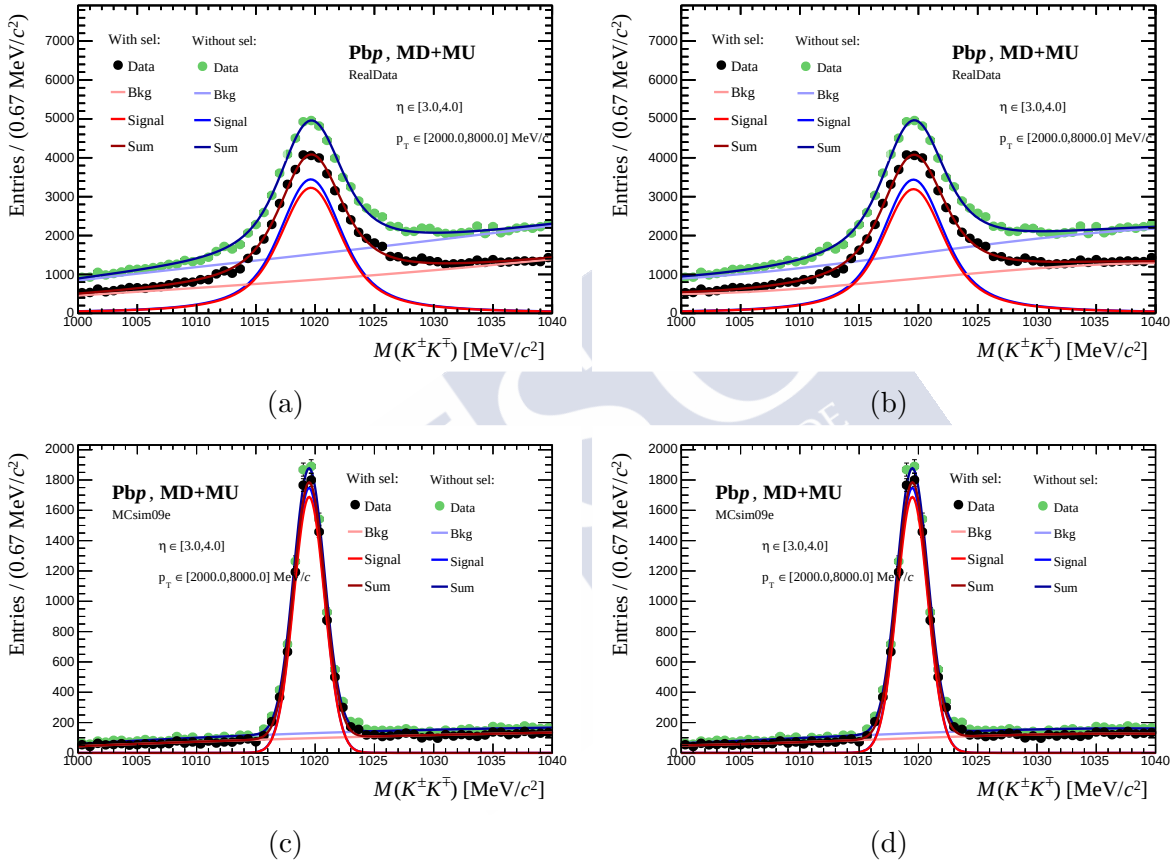


Figure 7.7: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ with the baseline (left) and alternative (right) background models. Data (top) and simulation (bottom) correspond to the $3 < \eta < 4$ and $2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$ kinematic bin of the Pbp configuration.

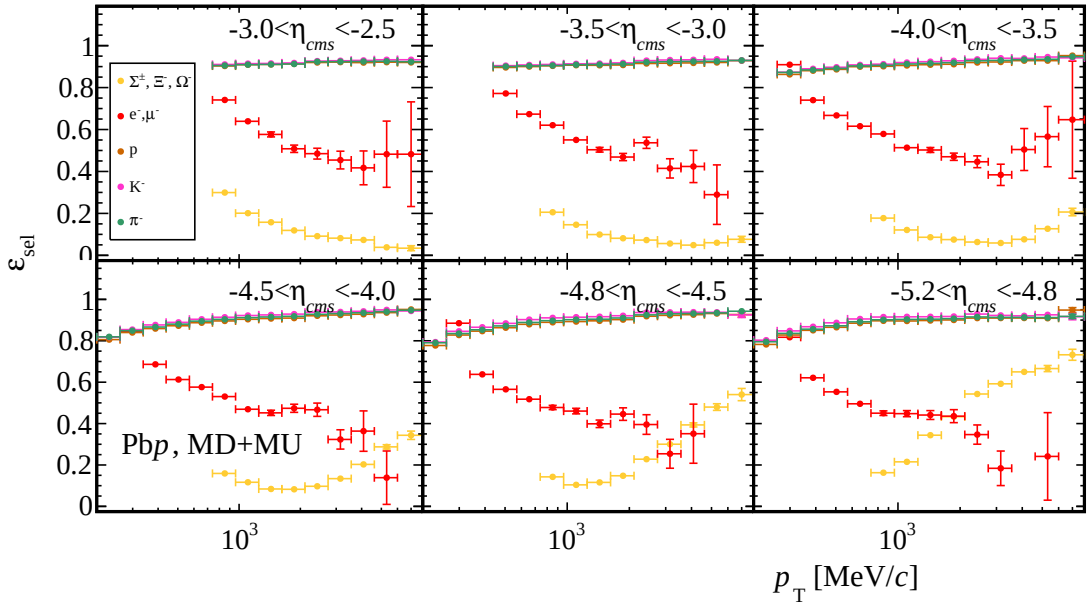


Figure 7.8: Variation of ε_{sel} with particle species for the Pbp configuration. In the legend, *cc*. for each particle must be inferred.

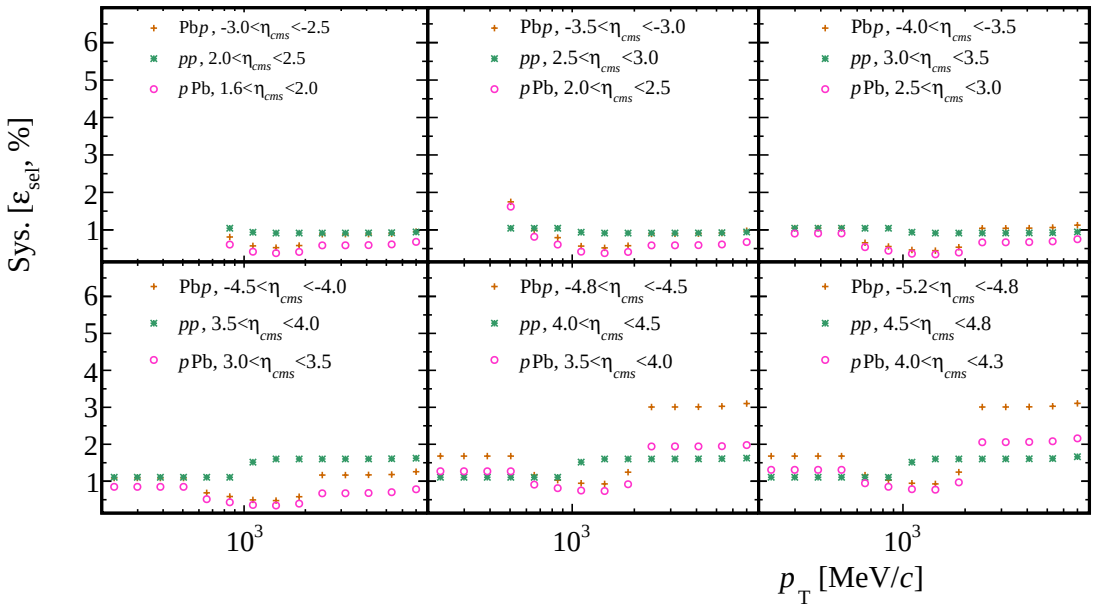


Figure 7.9: Relative systematic uncertainty in ε_{sel} for Pbp, pPb and pp data.

7.3 Background Subtraction

The systematic uncertainty of the background subtraction is determined by the systematic uncertainty of the different background components that appear in Eq. 6.22.

7.3.1 Fake tracks

The fake track fraction is given by Eq. 6.16: $f_{\text{fake}} = R_{\text{fake}} f^{\text{sim}}_{\text{fake}}$. The uncertainty of $f^{\text{sim}}_{\text{fake}}$, the simulation result, is statistical and originated in the limited size of the simulation sample and shown in the error bars of Fig. 6.19. The R_{fake} coefficient is calculated with Eq. 6.17 and is subject to systematic uncertainties arising from three sources:

1) The statistical uncertainty of $N^{\text{data}}_{\text{fake}}$ and $N^{\text{sim}}_{\text{fake}}$ in Eq. 6.17.

2) The purity of the proxy sample in data and simulation.

3) The arbitrary choice of the proxy sample.

While the uncertainty propagation of 1) to R_{fake} is straightforward the contributions of 2) and 3) require specific strategies.

The systematic uncertainty originated by 2) in R_{fake} is estimated as $(1-P^{\text{sim}}_{\text{fake}})R_{\text{fake}}$ for each (η, p_{T}) bin. Here $P^{\text{sim}}_{\text{fake}}$ is the fake track purity from simulation. This magnitude was already employed to optimise the proxy choice and is shown in Fig. 6.21 for the different samples. Since $P_{\text{fake}} > P^{\text{sim}}_{\text{fake}} \Rightarrow (1 - P_{\text{fake}})R_{\text{fake}} < (1 - P^{\text{sim}}_{\text{fake}})R_{\text{fake}}$, where P_{fake} is the fake track purity in data, the systematic uncertainty estimation is conservative.

The systematic uncertainty originated by 3) is estimated by considering alternative proxies for each beam configuration. These proxy samples are orthogonal to the nominal proxy samples and their definition is presented in Tab. 7.2. For pPb and Pbp shared VELO tracks with $0.1 <$ `GhostP` < 0.4 are used. For pp, non-shared VELO tracks are used along with the highest possible `GhostP` range, which is $0.375 <$ `GhostP` < 0.4, are considered. The purity for the alternative proxies is presented in Fig. 7.10. For Pbp and pPb the alternative proxy has a fake purity above 75% for all bins while for pp events lower purities are found in the $2 < \eta < 2.5$ range, as discussed in Sec. 6.4.1.

Statistical and the purity systematic uncertainty 2) are estimated in the alternative proxies R_{fake} whose results are shown in Fig. 7.11. In general, the trends are similar to the nominal proxy results of Fig. 6.22. The alternative result deviates from the nominal in the $2 < \eta < 2.5$ range for the pp sample, however the discrepancy is within the systematic uncertainty due to the poor fake purity of the alternative proxy.

An additional systematic uncertainty is considered to account for the discrepancy between the nominal and the alternative R results. The systematic uncertainty of 3) is computed as the weighted standard deviation of the nominal and the alternative R_{fake} values, taking as weights for the sum in quadrature of the systematic uncertainties from 1) and 2).

	pp	pPb	Pbp
Track type	Long	Long	Long
η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
p_{T}	$0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$	$0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$	$0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$
p	$p > 2\,\mathrm{GeV}/c$	$p > 2\,\mathrm{GeV}/c$	$p > 2\,\mathrm{GeV}/c$
`pseudoIP`	$< 0.368\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$	$< 0.348\,\mathrm{mm}$
`GhostP`	$0.375 <$ `GhostP` < 0.40	$0.1 <$ `GhostP` < 0.4	$0.1 <$ `GhostP` < 0.4
is shared VELO	`False`	`True`	`True`

Table 7.2: List of requirements for the alternative fake track proxy in each configuration. The standard (η, p_{T}) bins are considered.

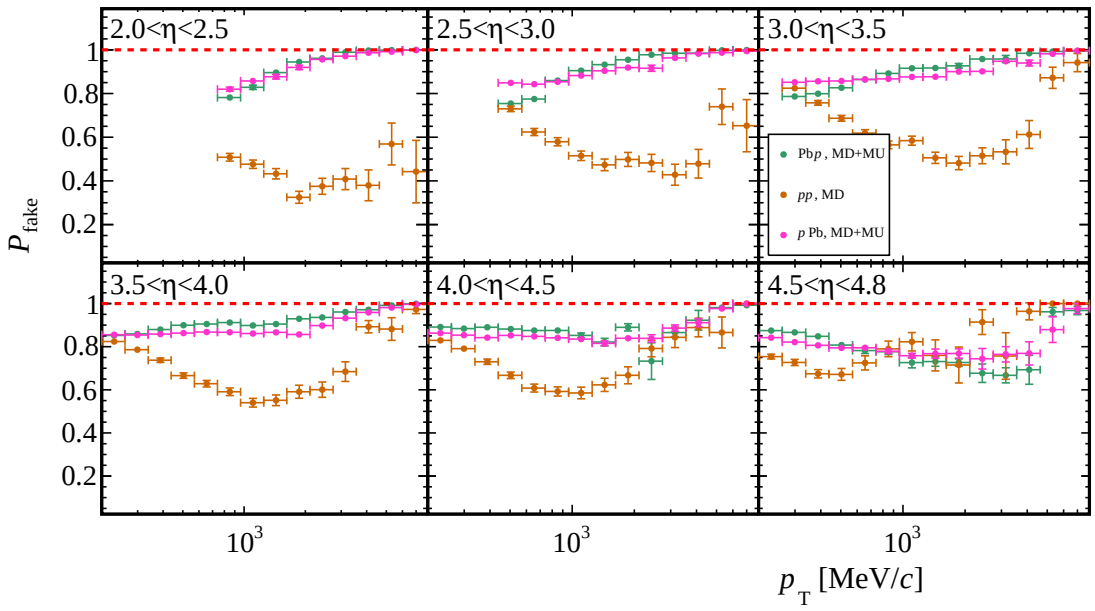


Figure 7.10: Fake track purity for the alternative proxies defined in Tab. 7.2 for pPb, Pbp and pp simulation samples. Error bars include statistical uncertainties.

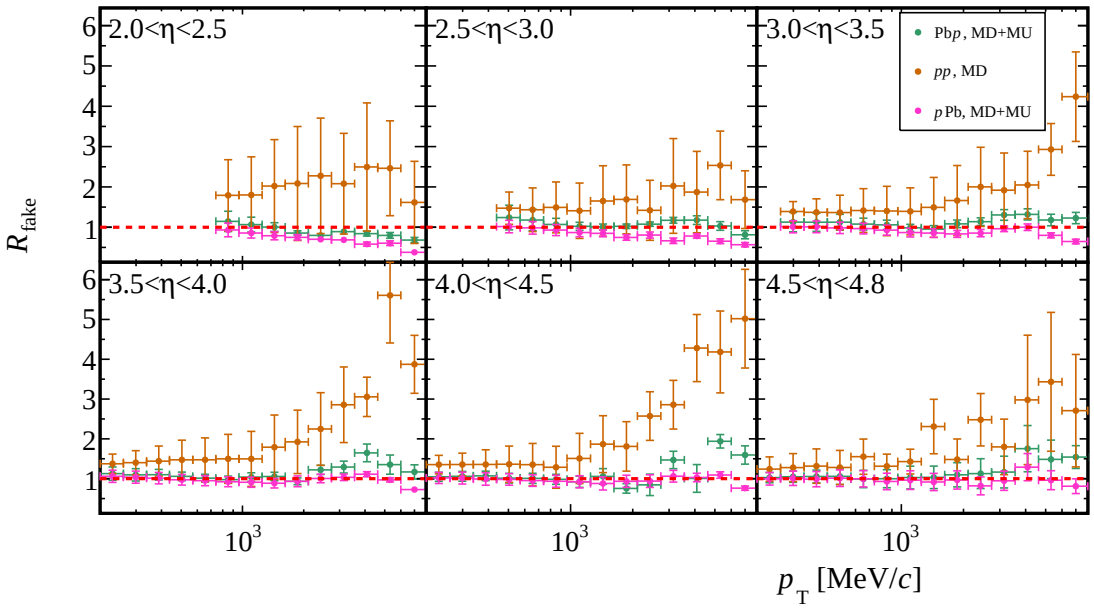


Figure 7.11: Discrepancy data over simulation R_{fake} for fake tracks in Pbp, pPb and pp data sets using the alternative proxy of Tab. 7.2. Simulation is weighted with set 1 for proton-led and set 12 for pp. Error bars include statistical and systematic uncertainty from the purity fake track purity in simulation.

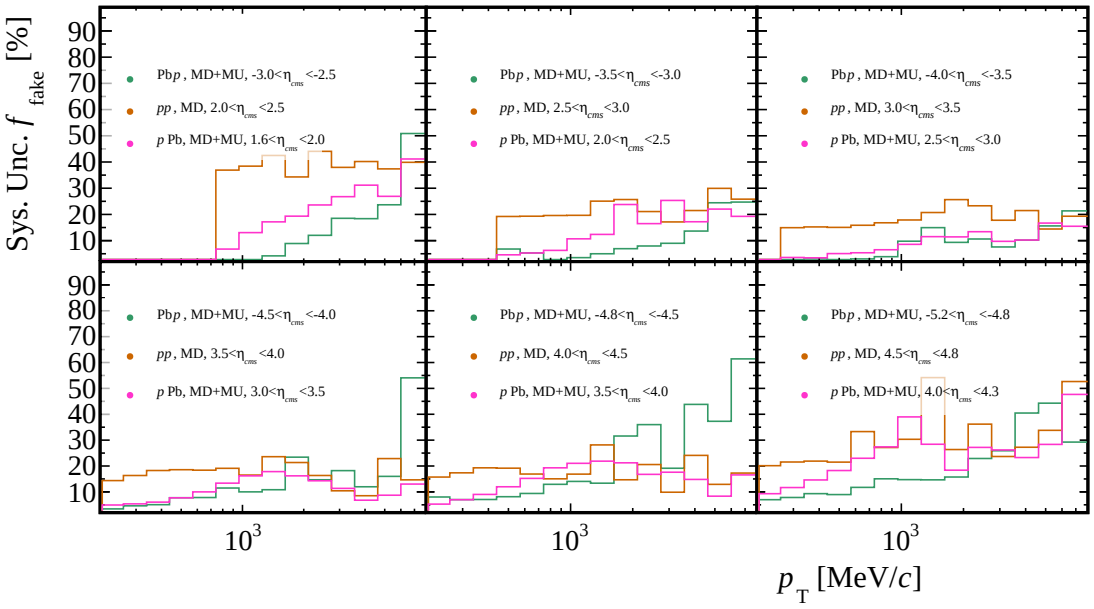


Figure 7.12: Relative systematic uncertainty on f_{fake} for Pbp, pPb and pp configurations.

The total R_{fake} uncertainty is estimated as the sum in quadrature of the effects given by 1), 2) and 3) and is shown in the error bars of Fig. 6.22. To understand the origin of the final uncertainty for each bin, the relative systematic uncertainty from each source is presented in Figs. B.23 and B.24. The uncertainty on R_{fake} is propagated together with the statistical uncertainty of $f_{\text{fake}}^{\text{sim}}$ to obtain the final uncertainty of f_{fake}, which is shown in the error bars of Fig. 6.23.

7.3.2 Secondary particles

The secondary particles fraction was discussed in Sec. 6.4.2, where a factorisation given by Eq. 6.19: $f_{\text{sec}} = R_\gamma f_\gamma^{\text{sim}} + R_{\text{had\&decays}} f_{\text{had\&decays}}^{\text{sim}} + R_{X\to\mu} f_{X\to\mu}^{\text{sim}}$ was introduced. The systematic uncertainties of R_γ, $R_{X\to\mu}$ and $R_{\text{had\&decays}}$, the factors that correct the simulation fractions of the different secondary types, are treated as for the R_{fake} case (see Sec. 7.3.1). The systematic uncertainty due to the contamination of the proxy sample is estimated as $(1-P_{\gamma,X\to\mu,\text{had\&decay}}^{\text{sim}})R_{\gamma,X\to\mu,\text{had\&decay}}$ for each (η, p_{T}) bin, being $P_{\gamma,X\to\mu,\text{had\&decay}}^{\text{sim}}$ the purity of the had&decay component from simulation. This systematic uncertainty is added in quadrature to the statistical uncertainty from the limited size of the proxy samples. Figs. 6.35, 6.36 and 6.37 show the value of $R_{\gamma,X\to\mu,\text{had\&decay}}$ with their systematic uncertainty.

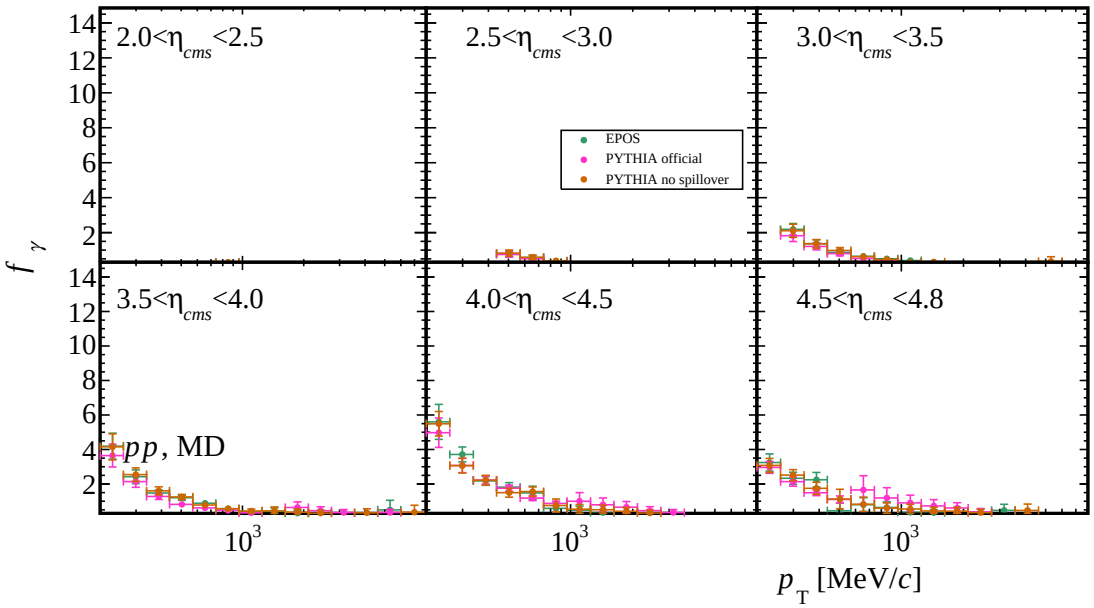
In addition, a cross-check of the procedure to obtain $f_{Had\&Decay}$ and f_γ for the *pp* configuration is performed using alternative simulation samples. These samples were privately generated with the following particularities:

- A sample of $1M$ events without spillover effects.
- A sample of $1M$ events generated with EPOS instead of Pythia and without spillover effects

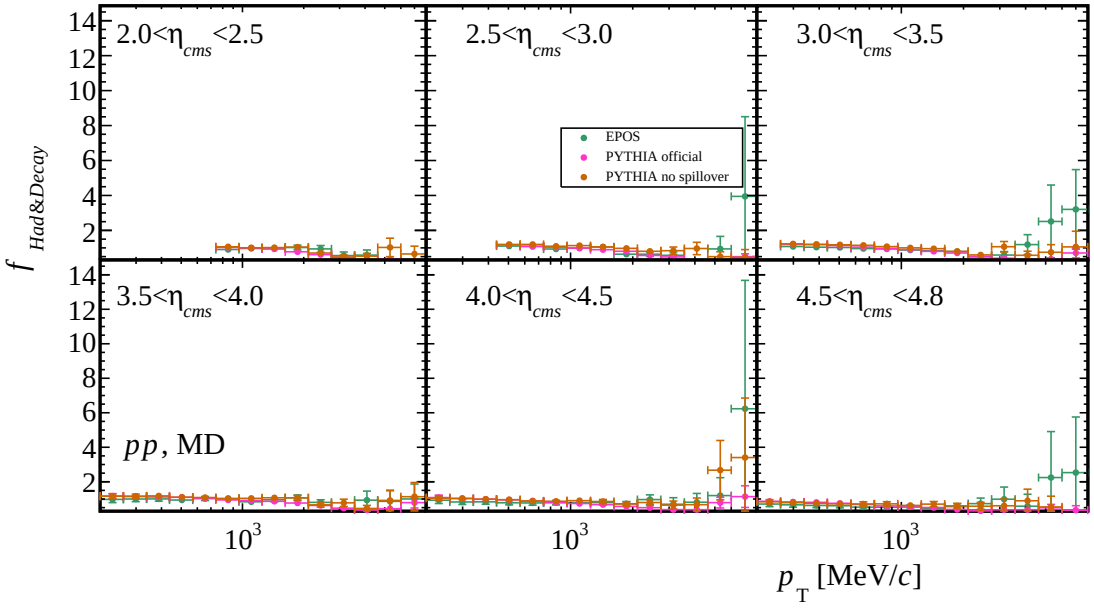
The samples are generated with Gauss following the same steps as the official sample apart from the above mentioned details.

The background from converted γs in the official *pp* simulation is overestimated in part because spill-over effects were included, although they are not present in data. By comparing f_γ computed with the two samples, which start from different values of f_γ^{sim}, the consistency of the procedure is verified. The EPOS sample allows to cross-check the estimation of $f_{\text{had\&decays}}$ because, as seen in Sec. 5.4.3, the Λ and K_{S}^0 have different abundances in EPOS than in Pythia. Fig. 7.13 presents f_γ and $f_{\text{had\&decays}}$ obtained with the nominal and the two alternative simulation samples. The three results show good agreement.

The systematic uncertainty in the fraction of secondary particles f_{sec} is obtained by propagating the uncertainties of each term in Eq. 6.19. The final relative systematic uncertainty on f_{sec} is shown in Fig. 7.14.



(a)



(b)

Figure 7.13: Comparison of f_γ (top) and $f_{\text{had\&decays}}$ (bottom) for pp estimated with official simulation and the two alternative simulation samples.

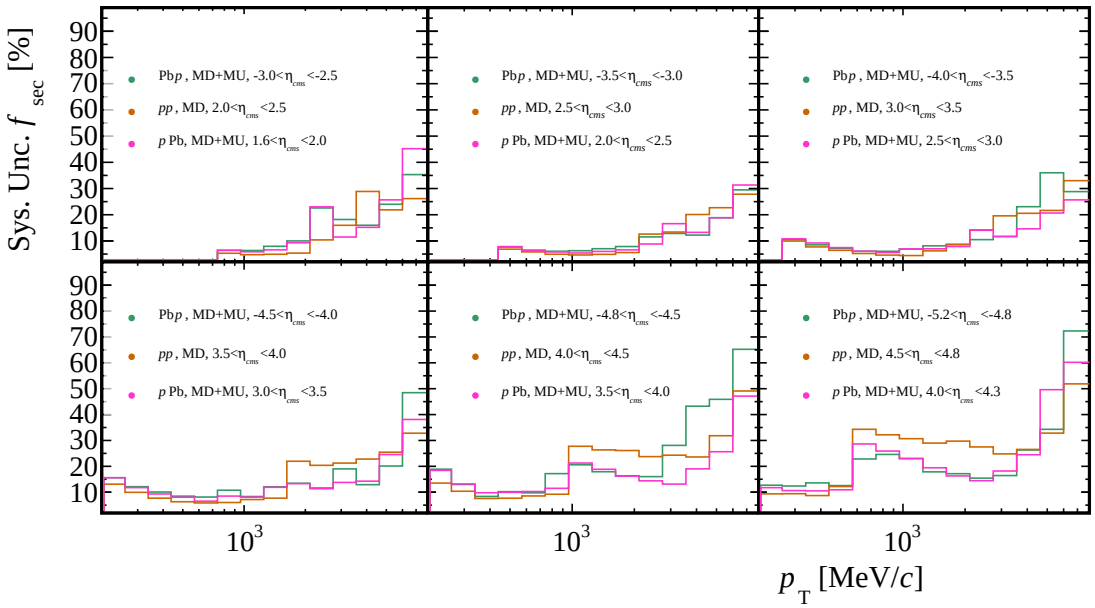


Figure 7.14: Relative systematic uncertainty on f_{sec} for Pbp, pPb and pp configurations.

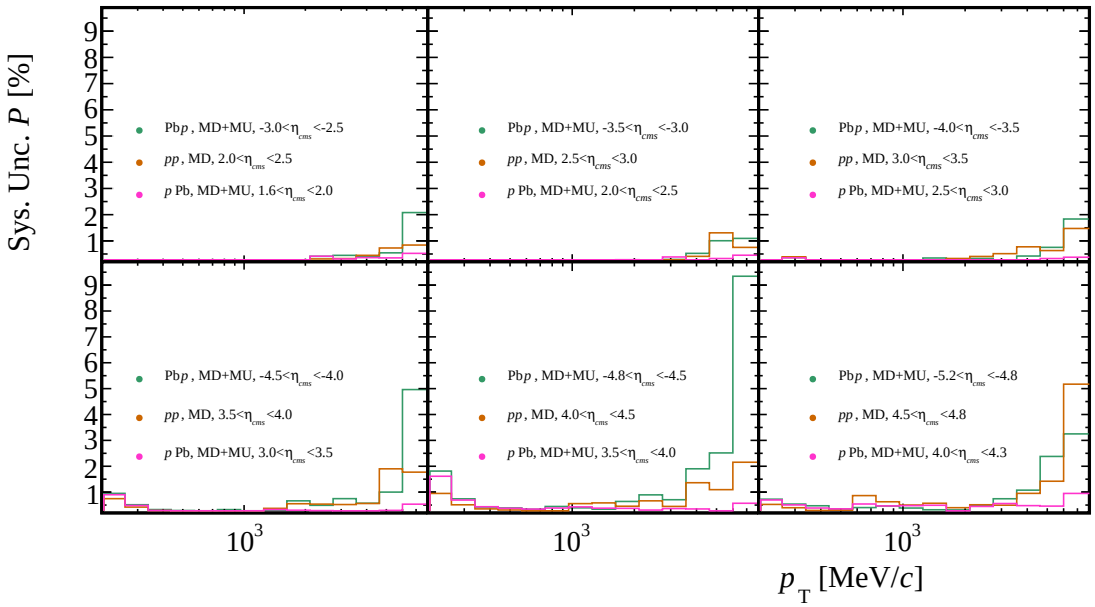


Figure 7.15: Relative systematic uncertainty on P for Pbp, pPb and pp configurations.

7.3.3 Summary

The final systematic uncertainty on P is calculated propagating the uncertainties of each factor in equation Eq. 6.22. All in all, the relative systematic uncertainty on P is presented in Fig. 7.15.

7.4 Candidate selection

In Eq. 6.1, N^{cand} depends on the track selection, principally in the (`pseudoIP`, `GhostP`) requirements. The values of $\varepsilon_{\text{reco}}$, ε_{sel}, P and N^{cand} allow to estimate N^{ch}. The consistency of the analysis can be cross-checked by evaluating the change in N^{ch} with modifications of the (`pseudoIP`, `GhostP`) criteria.

In this test, four different (`pseudoIP`, `GhostP`) pairs are used. They are shown in Tab. 7.3. The requirements are chosen varying the nominal values obtained in Sec. 5.3.3 by $\pm 10\%$. Fig. 7.16 shows how the alternative set of requirements are arranged with respect to the nominal set in the requirement optimisation. The values of ε_{sel}, P and ε_{TM} are calculated for each set. The reconstruction efficiency $\varepsilon_{\text{reco}}$ does not depend on the requirements on `GhostP` and `pseudoIP`.

Set	Pbp		pPb		pp	
	`pseudoIP` [mm]	`GhostP`	`pseudoIP` [mm]	`GhostP`	`pseudoIP` [mm]	`GhostP`
Nominal	0.348	0.109	0.348	0.103	0.368	0.078
1	0.29	0.109	0.29	0.103	0.31	0.078
2	0.348	0.12	0.348	0.12	0.368	0.10
3	0.43	0.109	0.43	0.103	0.43	0.078
4	0.348	0.095	0.348	0.09	0.368	0.063

Table 7.3: (`pseudoIP`, `GhostP`) pairs to verify the candidate selection consistency.

Figs. 7.18 and 7.17 show the ratio between N^{ch} obtained with each of the requirement sets and the nominal set. N^{ch} is calculated using Eq. 6.1. The variation of the yields with respect to nominal are $< 2\%$ for all bins in every configuration, and well within the systematic uncertainties.

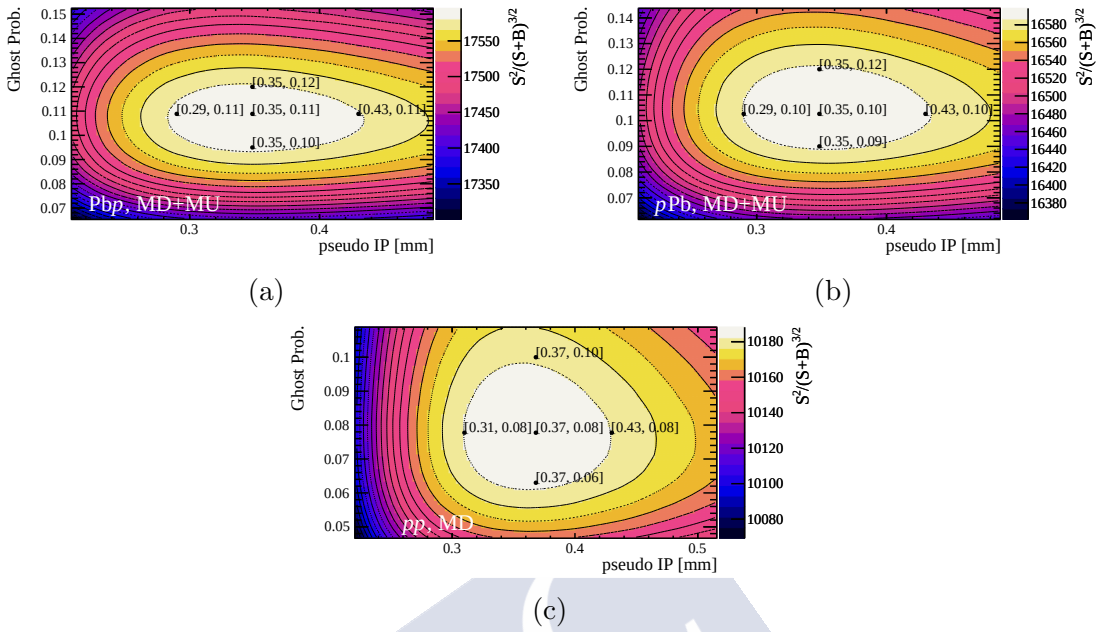


Figure 7.16: Distribution of the (`pseudoIP`, `GhostP`) requirements considered for the selection cross-check in the figure of merit of Eq. 5.8.

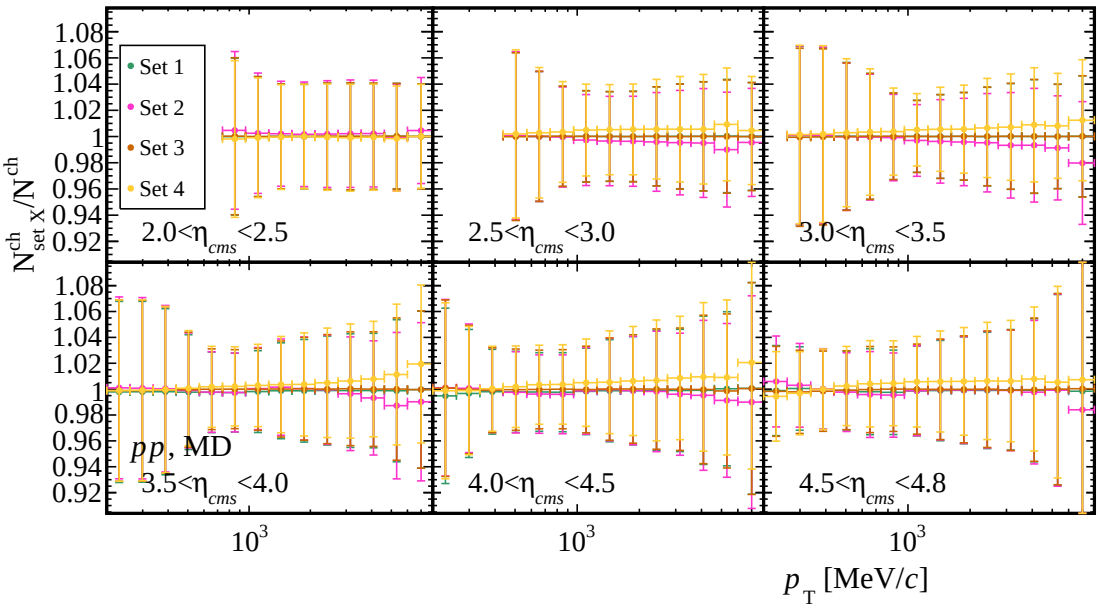
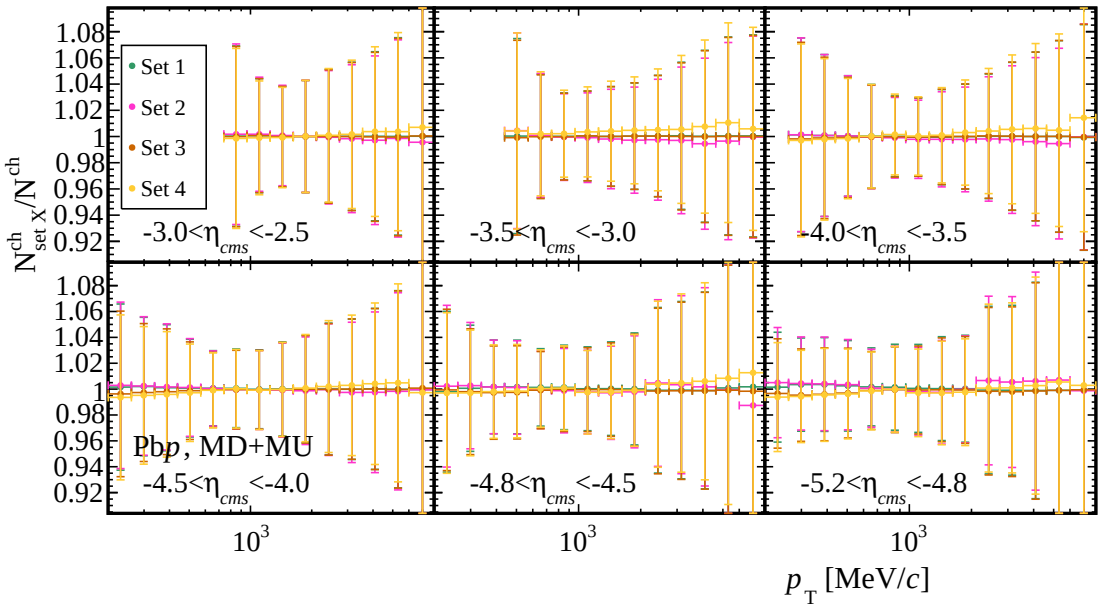
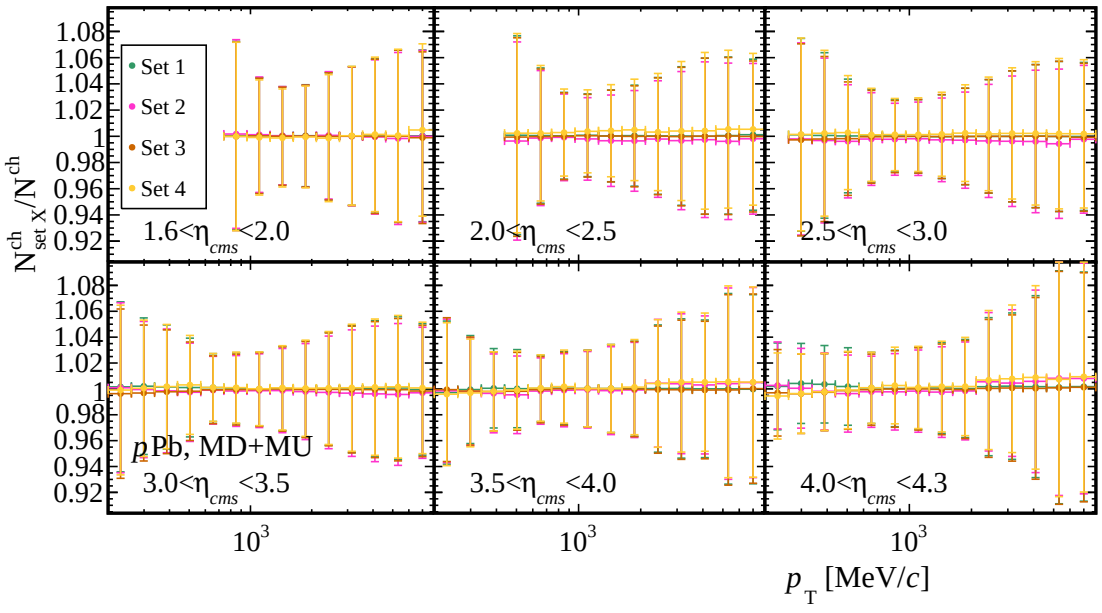


Figure 7.17: Ratio between the alternative N^{ch} with each set of requirements and the nominal result in pp. Error bars include statistical and systematic uncertainties.



(a)



(b)

Figure 7.18: Ratio between the alternative N^{ch} with each set of requirements and the nominal result in Pbp and pPb. Error bars include statistical and systematic uncertainties.

7.5 Other systematic uncertainties

7.5.1 Truth matching efficiency

A systematic uncertainty of the truth matching efficiency $\varepsilon_{\mathrm{TM}}$ is assigned from the yield errors of the fit procedure described in Sec. 6.5. This uncertainty is presented in Tab. 6.6.

7.5.2 Acceptance correction in proton-lead

The acceptance correction obtained in Sec. 6.6 is estimated with simulation. Since the correction depends on the particle composition, an effect arises from the unknown particle composition in data. However, as presented in Fig. 6.45, the correction is about 0.5% for low p_{T} and negligible for high p_{T} bins. As it was seen in Secs. 5.4.3 and 7.1.3, low p_{T} bins are mainly pion-dominated and the contribution from kaons and protons only becomes relevant for $\varepsilon_{\mathrm{reco}}$ at high p_{T}. Therefore, the systematic uncertainty due to the unknown particle composition in the acceptance correction is negligible.

7.5.3 Bin migration

The unfolding cross-check in Sec. 6.7 proves that bin migration effect does not have a significant effect given the LHCb resolution and the proposed bin size. The largest effect occurs in the lowest p_{T} bin and is about 0.5%. However, as it was seen previously, the reconstruction efficiency already accounts for bin migration, and therefore it is not necessary to assign a systematic uncertainty.

The bin migration study was performed using simulation samples, and therefore relies on the correct description of detector resolution by simulation. The resolution agreement between data and simulation can be verified by comparing the σ parameter of the Voigtian function in the signal extraction of $\phi \to K^{\pm}K^{\mp}$ described in Sec. 6.3.1. Tab. 7.4 shows the results of σ for each (p_{T}, η) bin in Pbp, pPb and pp. A good level of agreement between data and simulation resolutions is observed.

p_{T} [GeV/c]; η	σ, pPb		σ, Pbp	
	Data	MC	Data	MC
$0.2 < p_{\mathrm{T}} < 0.6; 2 < \eta < 3$	0.964 ± 0.072	0.803 ± 0.015	0.674 ± 0.097	0.809 ± 0.018
$0.2 < p_{\mathrm{T}} < 0.6; 3 < \eta < 4$	1.067 ± 0.034	1.034 ± 0.010	1.100 ± 0.041	1.038 ± 0.012
$0.2 < p_{\mathrm{T}} < 0.6; 4 < \eta < 4.8$	1.544 ± 0.041	1.451 ± 0.022	1.580 ± 0.050	1.487 ± 0.026
$0.6 < p_{\mathrm{T}} < 1; 2 < \eta < 3$	0.929 ± 0.031	0.8453 ± 0.0068	0.948 ± 0.039	0.8773 ± 0.0079
$0.6 < p_{\mathrm{T}} < 1; 3 < \eta < 4$	1.243 ± 0.018	1.1141 ± 0.0068	1.236 ± 0.022	1.1431 ± 0.0078
$0.6 < p_{\mathrm{T}} < 1; 4 < \eta < 4.8$	1.664 ± 0.030	1.587 ± 0.017	1.657 ± 0.035	1.634 ± 0.019
$1 < p_{\mathrm{T}} < 2; 2 < \eta < 3$	0.915 ± 0.022	0.8847 ± 0.0052	1.009 ± 0.025	0.9072 ± 0.0058
$1 < p_{\mathrm{T}} < 2; 3 < \eta < 4$	1.325 ± 0.016	1.2029 ± 0.0065	1.301 ± 0.020	1.2035 ± 0.0066
$1 < p_{\mathrm{T}} < 2; 4 < \eta < 4.8$	1.765 ± 0.033	1.629 ± 0.017	1.703 ± 0.038	1.694 ± 0.018
$2 < p_{\mathrm{T}} < 8; 2 < \eta < 3$	1.185 ± 0.034	1.006 ± 0.011	1.131 ± 0.044	1.005 ± 0.011
$2 < p_{\mathrm{T}} < 8; 3 < \eta < 4$	1.359 ± 0.037	1.306 ± 0.015	1.488 ± 0.050	1.297 ± 0.016
$2 < p_{\mathrm{T}} < 8; 4 < \eta < 4.8$	1.80 ± 0.10	1.685 ± 0.046	1.55 ± 0.15	1.608 ± 0.051

p_{T} [GeV/c]; η	σ, pp	
	Data	MC
$0.2 < p_{\mathrm{T}} < 1; 2 < \eta < 3.5$	1.039 ± 0.014	0.940 ± 0.066
$0.2 < p_{\mathrm{T}} < 1; 3.5 < \eta < 4.8$	1.380 ± 0.041	1.349 ± 0.059
$1 < p_{\mathrm{T}} < 8; 2 < \eta < 3.5$	1.095 ± 0.015	0.892 ± 0.072
$1 < p_{\mathrm{T}} < 8; 3.5 < \eta < 4.8$	1.491 ± 0.022	1.600 ± 0.094

Table 7.4: Result for the σ parameter of the Voigtian function in the signal extraction of $\phi \to K^{\pm}K^{\mp}$.

7.6 Summary of the uncertainties

In Tab. 7.5, a summary of the relative uncertainty from every considered source as well as the total uncertainty in the different observables are presented. The total uncertainty in the differential cross-sections is between 2.8% and 14.5%, and the variation with $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$, shown in Fig. 7.19, follows similar trends for pPb, Pbp and pp. Most bins have an uncertainty below 5%.

For more illustration, Figs. 7.20, 7.21 and 7.22 show the relative contribution to the total uncertainty per source in Pbp, pPb and pp, respectively. In very low p bins, the uncertainty is dominated by the C_{tracking} uncertainty, motivated by the lack of statistics in the calibration sample of $J/\psi \to \mu^+\mu^-$ for low momentum tracks. It is worth noting that tracks without tracking correction contribute with a 5% to the systematic uncertainty, while for the others the contribution is typically below 2%. At higher momentum, the measurement is dominated by the luminosity uncertainty of 2.3%, 2.5% and 2.0% for pPb, Pbp and pp, respectively. The bins dominated by the luminosity uncertainty correspond to those with the highest precision measurement. At higher p and p_{T}, the uncertainty starts to become dominated by the particle composition. This is because, as explained in Sec. 7.1.3, the relative abundance of kaons, protons and hyperons with respect to pions rises with p_{T} thus the average reconstruction efficiency becomes more sensitive to

Uncertainty Source	pPb [%]	Pbp [%]	pp [%]
$C_{tracking}$	1.5-5.0	1.5-5.0	1.6-5.3
Occupancy	0.0-2.8	0.6-2.9	0.1-1.6
Part. Composition	0.4-4.1	0.4-4.6	0.3-2.4
$\varepsilon_{\mathrm{sel}}$	0.7-2.2	0.7-3.0	1.0-1.7
P	0.1-1.8	0.1-11.7	0.1-5.8
$\varepsilon_{\mathrm{TM}}$	0.0-0.1	0.0-0.1	0.1-0.2
$\mathcal{L}^{\dagger}$	2.3	2.5	2.0
Statistical	0.0-0.6	0.0-1.0	0.0-1.1
Total (in $d^2\sigma/d\eta_{\mathrm{cms}}dp_{\mathrm{T}}$)	3.0-6.7	3.3-14.5	2.8-8.7
Total (in $R_{p\mathrm{Pb}}$)	4.2-9.2	4.4-16.9	
Total (in R_{BF})	4.5-11.8		

Table 7.5: Relative uncertainties for Pbp, pPb and pp. The interval indicates the minimum and the maximum relative value among all the $(\eta_{\mathrm{cms}}, p_{\mathrm{T}})$ bins. The † symbol denotes a fully correlated uncertainty.

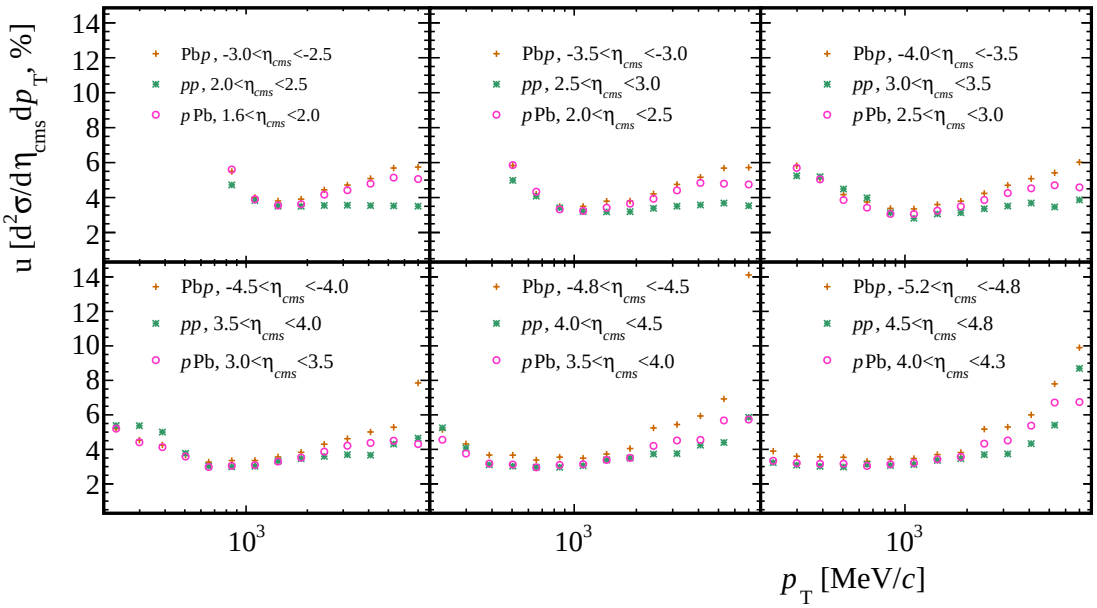


Figure 7.19: Total relative uncertainty in $d^2\sigma/d\eta dp_{\mathrm{T}}$ for Pbp, pPb and pp.

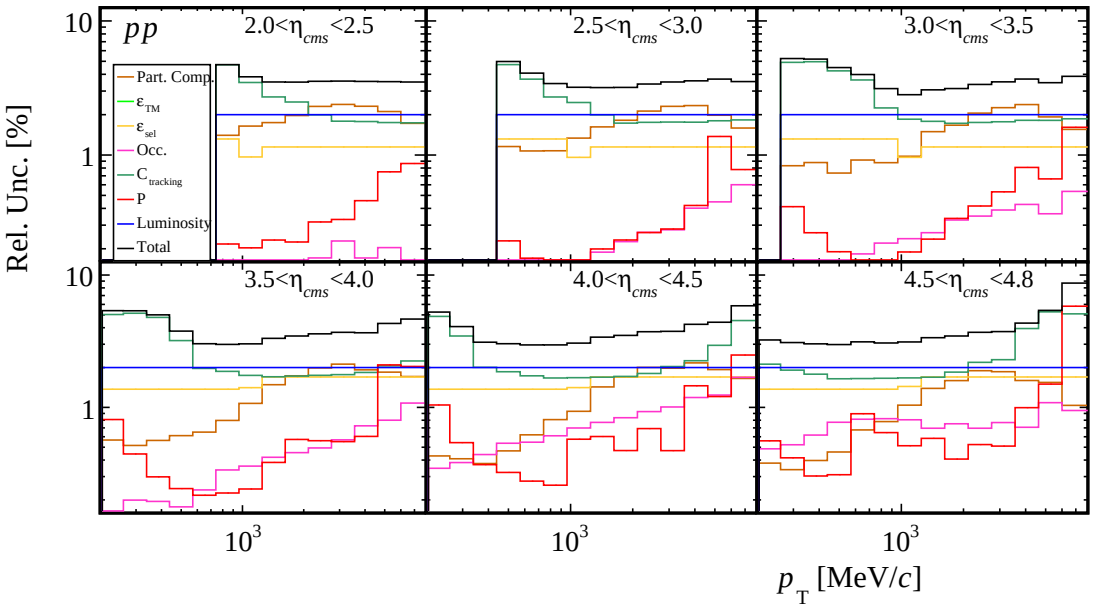


Figure 7.20: Relative uncertainty per source in the pp cross-section measurement with respect to η and p_{T}. The statistical uncertainty is not shown.

variations in the single-particle efficiencies. The effect is more pronounced in Pbp and then in pPb, as strangeness and baryon production are enhanced in those systems with respect to pp [93]. For pp the effect is smaller and this contribution is at the level of the luminosity uncertainty.

For very high p bins, the C_{tracking} contribution dominates again because of the lack of statistics of the calibration sample for those tracks. This effect is specially seen in the most forward (backward for Pbp) η_{cms} bin. Finally, for the Pbp sample, the P contribution dominates for two of the last p_{T} bins ($-4.5 < \eta_{\mathrm{cms}} < -4.0$ and $-4.8 < \eta_{cms} < -4.5$). The origin of this increase is in the uncertainty in the fake track background estimation, as can be checked in Figs. 6.19, 7.12 and 7.15. It can be observed, specially in Fig. 6.19, that $f_{\mathrm{fake}}^{\mathrm{sim}}$ has a large uncertainty for those bins that is then propagated to the purity. The reason is a small fake track statistics in the reweighted simulation sample in those intervals. This is related to large weights in high multiplicity events to compensate the discrepancy between data and simulation, see Fig. 5.5, and the strong dependence of $f_{\mathrm{fake}}^{\mathrm{sim}}$ with detector occupancy shown in Figs. 6.17a and 6.17b.

With this situation, the systematic uncertainties could be reduced in future analyses. First, an important reduction of systematic uncertainties can be achieved by measuring the abundances of π, K, p and hyperons in proton-lead and proton-proton collisions. The measured relative abundances can be then used as input to Eq. 6.8. The systematic effect from the particle composition would then depend on the achieved experimental uncertainty, but a great improvement with respect to the 30% coverage factor applied to

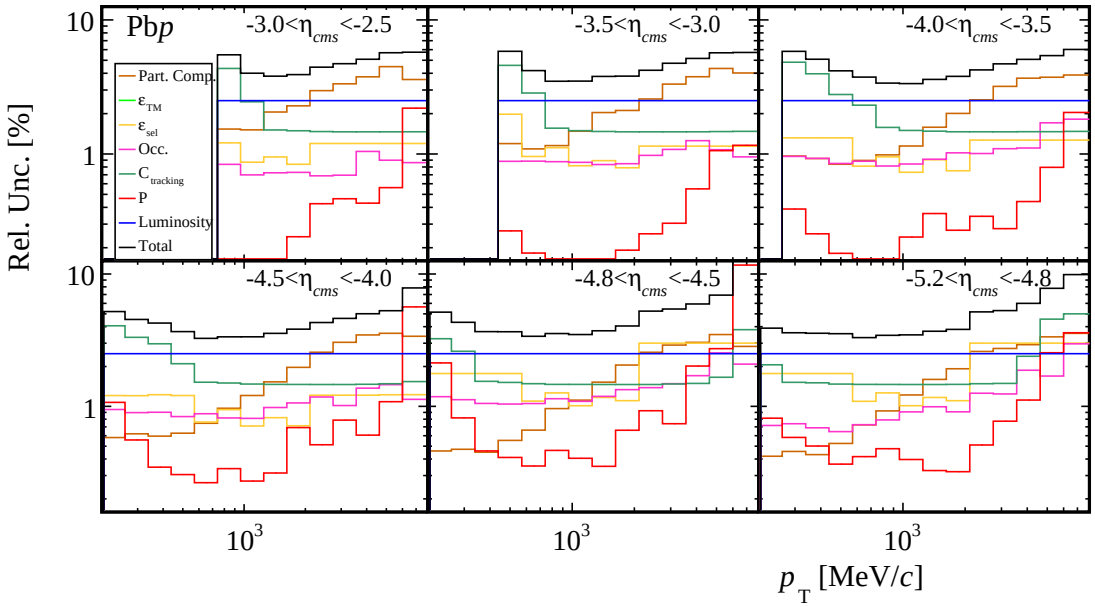


Figure 7.21: Relative uncertainty per source in the Pbp cross-section measurement with respect to η_{cms} and p_{T}. The statistical uncertainty is not shown.

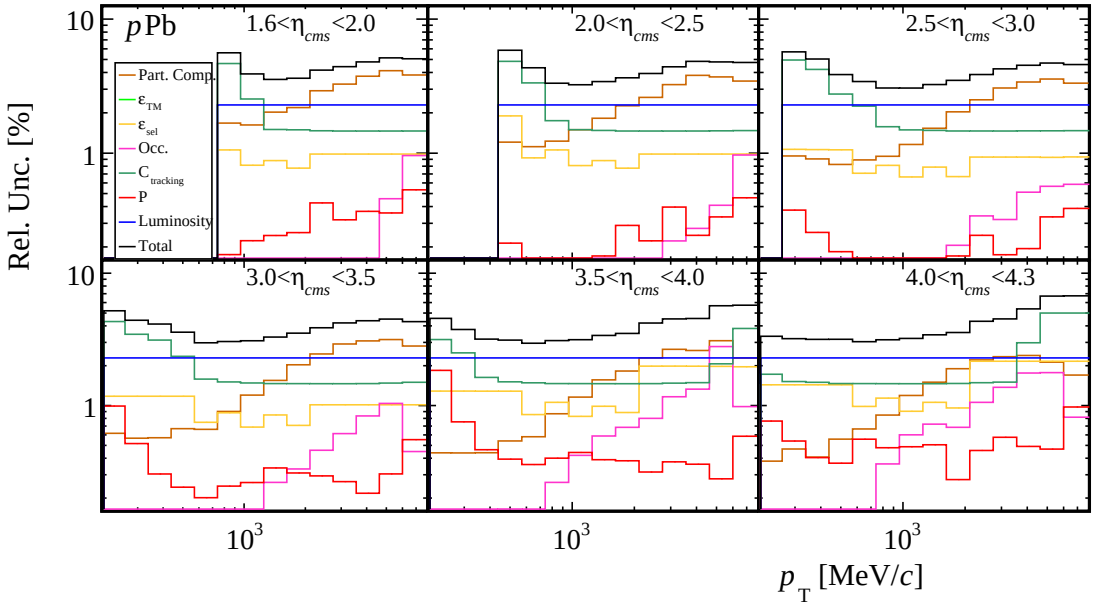


Figure 7.22: Relative uncertainty per source in the pPb cross-section measurement with respect to η and p_{T}. The statistical uncertainty is not shown.

the current relative abundances is expected.

To reduce the systematic uncertainty from C_{tracking}, additional calibration samples could be prepared and considered. The decay channels $K^0_S \to \pi^+\pi^-$ and $Z \to \mu^+\mu^-$ might be adequate to increase the coverage for low p tracks and for high p tracks respectively. For the last case, the luminosity of the 2013 pPb and Pbp samples might not be sufficient to have large enough $Z \to \mu^+\mu^-$ samples for an appropriate calibration sample. Finally, the systematic uncertainty from P can be reduced by increasing the sample of high p_T fake tracks in high multiplicity events in the Pbp simulation sample. An important improvement could be achieved with a simulation sample with similar size to the current one but with occupancy distributions closer to data. The mentioned improvements are also mandatory to extend the measurement to a higher p_T range.

8

Results and discussion

In this chapter, the results of the analysis are presented and discussed. The double-differential cross-section is shown in Sec. 8.1, and the result for pp collisions is compared to other measurements in Sec. 8.1.1. Then, the nuclear modification factor, the forward-to-backward ratio and the backward-to-forward ratio are presented in Sec. 8.2. As for the cross-sections, the results are compared with other measurements from other experiments at LHC and RHIC in Sec. 8.2.1. Also, comparisons with predictions from selected phenomenological models are made in Sec. 8.2.2. Finally, the dependence of the nuclear modification with the x and Q^2 variables is studied in Sec. 8.2.3. In the following, the pseudorapidity in the nucleon-nucleon centre-of-mass system, previously η_{cms}, is referred to as η to simplify the notation.

8.1 Double-differential cross-section

The double-differential cross-sections of prompt charged particle production can be obtained from the prompt charged particle yields $N^{\mathrm{ch}}(\eta, p_{\mathrm{T}})$ and the integrated luminosity of the datasets as shown in Eq. 2.12. The $N^{\mathrm{ch}}(\eta, p_{\mathrm{T}})$ yields are directly obtained with Eq. 6.1 with the $N_{\mathrm{candidates}}$ distributions from the data samples and the $\varepsilon_{\mathrm{total}}$ values presented in Fig. 6.51. The integrated luminosity of the considered datasets is known, as explained in Secs. 4.1.1 and 4.2.1.

The production cross-section of pPb, Pbp and pp collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ are shown in Fig. 8.1. The error bars account for statistical, systematic and luminosity

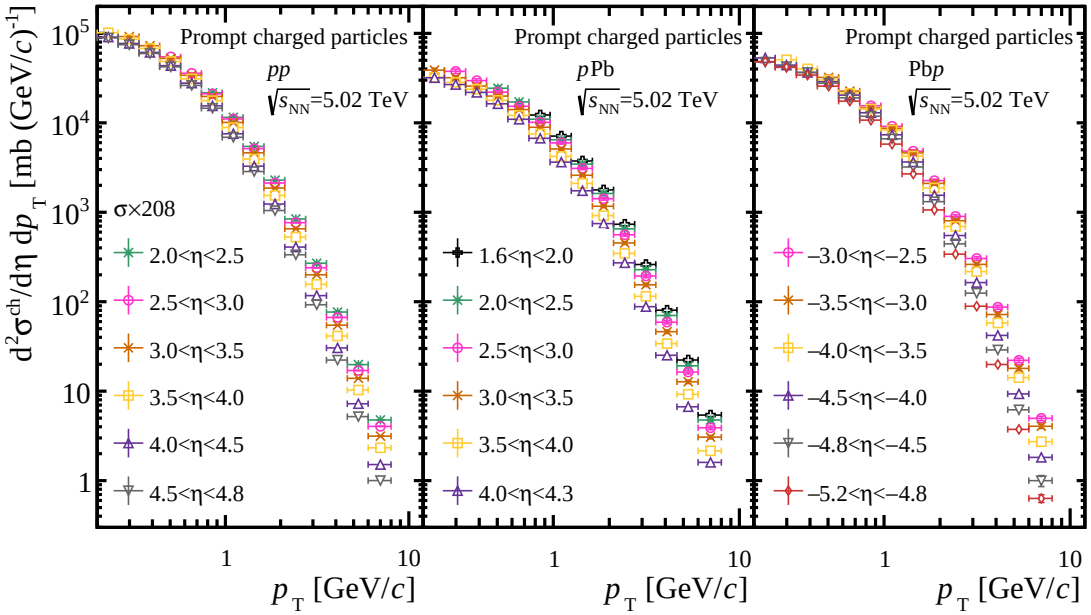


Figure 8.1: Charged particle production cross-section in Pbp (top left), pPb (top right) and pp (bottom) collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. Error bars account for statistical, systematic and luminosity uncertainties added in quadrature.

uncertainties added in quadrature. To ease the comparison between pp, pPb and Pbp, the pp result is multiplied by $A = 208$, the number of nucleons in the lead nuclei.

The differential cross-section can be interpreted as the average number of prompt charged particles produced in a given η and p_{T} range in a pPb (pp) collision at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ multiplied by the inelastic pPb or pp cross-section at the same energy. The inelastic cross-section for pPb and pp collisions has not been measured at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ yet, so the estimated average of charged particles in a given collision is not directly accessible. Instead, the total inelastic nucleon-nucleon cross-section of $67.6 \pm 0.6\,\mathrm{mb}$ at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$, which is obtained by interpolation [177], can be used to scale the cross-section results. This implies on average 0.1035 ± 0.0029 (10.87 ± 0.34) prompt charged particles produced in a pp (pPb) collision, with $0.961 < p_{\mathrm{T}} < 1.249\,\mathrm{GeV}/c$ and $3.0 < \eta < 3.5$. For $-3.5 < \eta < -3.0$ in Pbp collisions, the average number of charged particles would be 18.53 ± 0.65 for the same p_{T} range.

Fig. 8.2 shows the cross-section for Pbp and pp events with an alternative η_{cms} binning scheme for some η values, which is necessary to compute the $R_{p\mathrm{Pb}}$ and the R_{FB} observables in all the available acceptance. This follows from the discussion in Sec. 6.1.

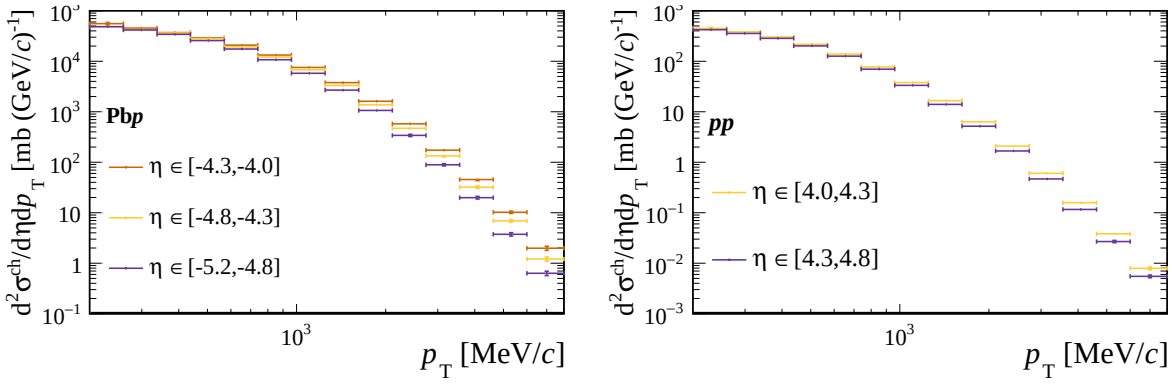


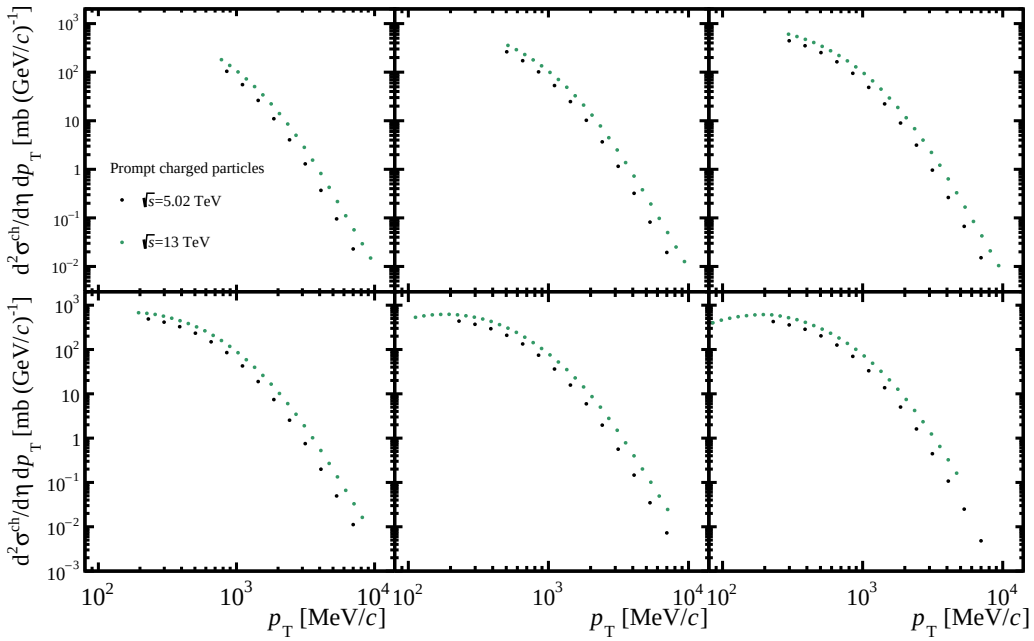
Figure 8.2: Charged particle production cross-section in Pbp (left) and pp (right) collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ with the alternative binning as discussed in Sec. 6.1. Error bars account for statistical, systematic and luminosity uncertainties added in quadrature.

8.1.1 Comparison with other experiments

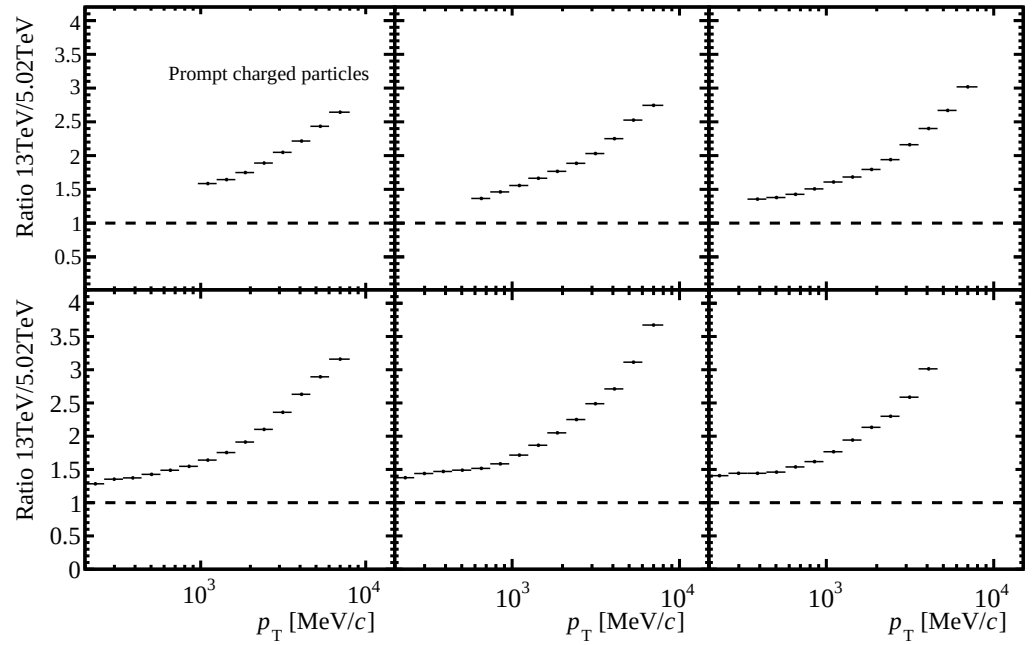
Although the prompt charged particle production cross-section for pp nor pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ were measured before, some comparisons can be made with results at different collision energies and different kinematic ranges. The comparisons are limited to the pp result. For pPb and Pbp, the previous determinations of charged particle production at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ [41, 42, 69] determined the invariant yields, which are not comparable to the cross-sections unless the total inelastic cross-section is known.

The pp cross-sections at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ can be compared with a recent LHCb measurement at $\sqrt{s_{\mathrm{NN}}} = 13\,\mathrm{TeV}$ [178]. Both cross-sections are shown in Fig. 8.3a, where no uncertainties are included in the plot. Appart from the differences in the binning and the p_{T} range, the production of charged particles grows with the collision energy as previously seen by the ALICE experiment in the central pseudorapidity region [41, 179]. To study the p_{T} trend, the 13 TeV result is interpolated using the ROOT `TSpline3` class [164] to the binning of the 5.02 TeV measurement, and the ratio between both results is computed and shown in Fig. 8.3b, where the uncertainties are not considered to avoid addressing the effect from the different binning. The ratio increases with p_{T}, similarly to the measurements from ALICE [179] in the central region.

Additionally, the pp result can be compared with the ALICE measurement at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ which covers $|\eta| < 0.8$ [41]. Both results are presented in Fig. 8.4a, showing a comprarable particle production rate in the ALICE acceptance and in the $\eta \in [2.0, 2.5]$ bin of this work. To put this results in context the charged particle production with respect to pseudorapidity in pp collisions at $\sqrt{s_{\mathrm{NN}}} = 8\,\mathrm{TeV}$ is shown as measured by the CMS and TOTEM collaborations [180]. The measurement covers $|\eta| < 2.2$ and $5.3 < |\eta| < 6.4$, which does not match the LHCb acceptance, but indicates a smooth variation of charged particle production in $0 < \eta < 6.4$. This is consistent with the observed behaviour of the cross-sections of LHCb and ALICE in Fig. 8.4a.

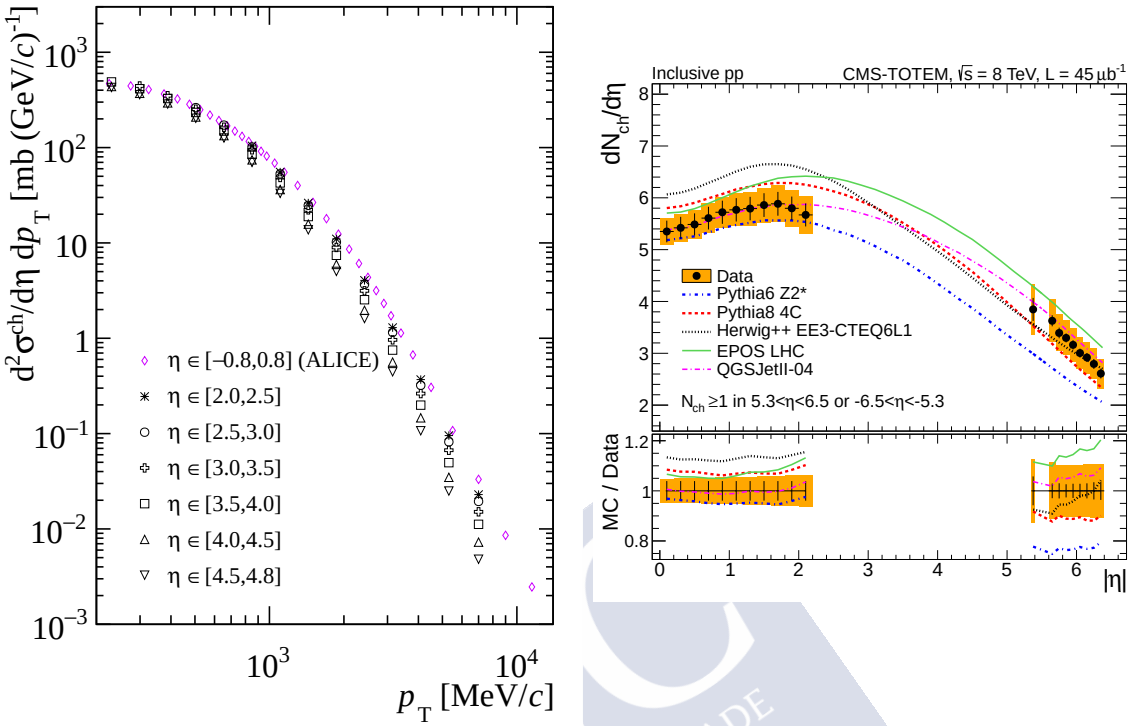


(a)



(b)

Figure 8.3: Fig. 8.3a: Comparison of the LHCb prompt charged particle cross-section at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ (this work) and $\sqrt{s_{\rm NN}} = 13\,\text{TeV}$ [178] for pp collisions. Fig. 8.3b: Ratio between both cross-sections. Uncertainties are not shown in any of the figures since the goal is a qualitative comparison between both results.



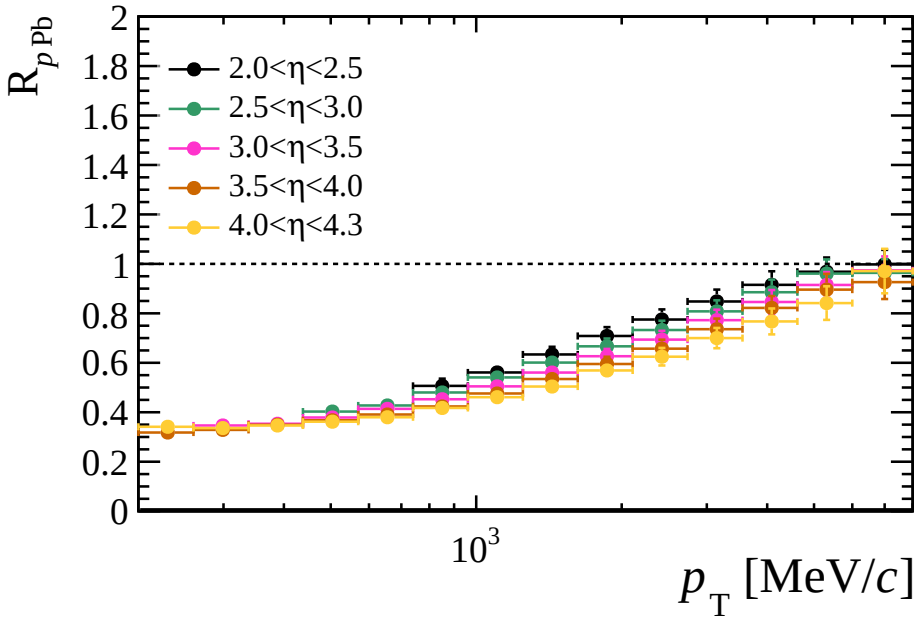
(a) (b)

Figure 8.4: Fig. 8.4a: comparison of the prompt charged particle cross-section for pp collisions at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ in the forward region (this work) and in the central region (ALICE measurement [41]). Fig. 8.4b: pseudorapidity dependent densities of charged particles in pp collisions as measured by the CMS and TOTEM collaborations [180], and compared by predictions of several models.

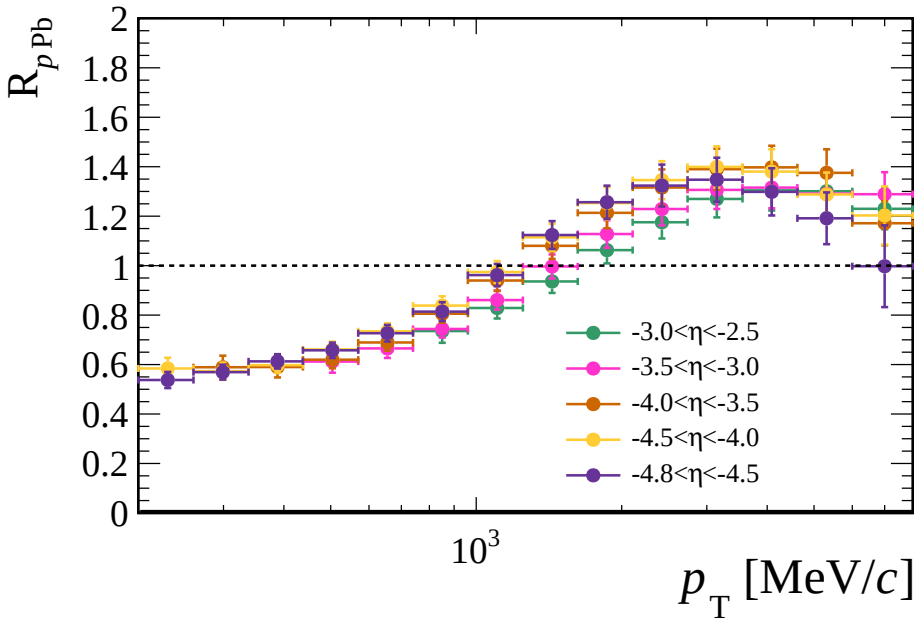
8.2 Nuclear modification factor

The nuclear modification factor $R_{p\text{Pb}}$ is obtained with Eq. 2.11 and the cross-sections presented in Fig. 8.1 and Fig. 8.2. The results for $R_{p\text{Pb}}$ in different $(\eta, p_{\rm T})$ intervals are presented in Fig. 8.5a and Fig. 8.5b for the forward and backward regions respectively. The uncertainties account for statistical, systematic and luminosity sources added in quadrature.

In the forward region (Fig. 8.5a), the measurement indicates a suppression of charged particle production in proton-lead collisions relative to that in scaled pp collisions, which increases towards forward pseudo-rapidities. In the low $p_{\rm T}$ regime, $R_{p\text{Pb}}$ reaches values of about 0.3 in the most forward pseudo-rapidities. There is a significant difference between the $R_{p\text{Pb}}$ value at $2.0 < \eta < 2.5$ and $4.0 < \eta < 4.3$ around $1 \gtrsim p_{\rm T} \gtrsim 3\,\text{GeV}/c$. At the



(a)



(b)

Figure 8.5: Nuclear modification factor for the forward and backward region. Error bars account for statistical, systematic and luminosity uncertainties added in quadrature.

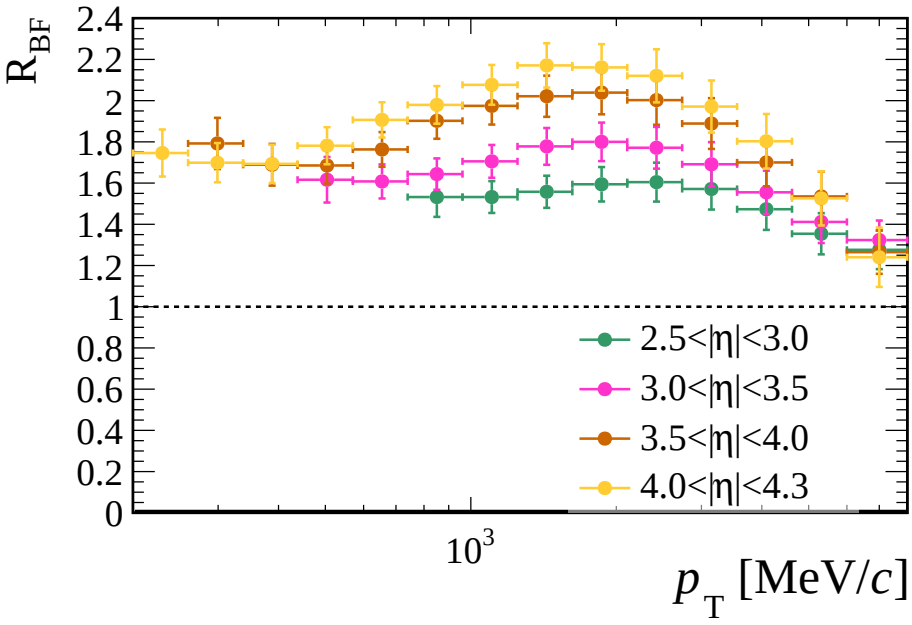


Figure 8.6: Backward-to-forward ratios. Error bars account for statistical, systematic and luminosity uncertainties added in quadrature.

highest p_{T} ($6 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$), the result is compatible with no nuclear modification for all η intervals.

In the backward region a significant enhancement of charged particle production is observed for $p_{\mathrm{T}} > 1.5\,\mathrm{GeV}/c$ for all η ranges. Similar enhancements have been interpreted as a Cronin-like effect [181], as discussed in Sec. 2.4.3. The enhancement reaches a maximum at different p_{T} values depending on η, followed by a decreasing trend towards unity. This decrease is more pronounced in the most backward pseudorapidity. The initial increasing trend starts at lower p_{T} values at the most backward pseudorapidity with respect to the less backward region. The maximum value of $R_{p\mathrm{Pb}}$ is found to be ~ 1.3 and depends slightly on η. The value of $R_{p\mathrm{Pb}}$ at low p_{T}, $R_{p\mathrm{Pb}} \approx 0.6$, is about a factor two higher than the value observed in the forward region, $R_{p\mathrm{Pb}} \approx 0.3$.

The backward-to-forward ratio R_{BF} is obtained with Eq. 2.13 and the cross-sections shown in Fig. 8.1 and Fig. 8.2. The results for R_{BF} in different (η, p_{T}) intervals are presented in Fig. 8.6. For every (η, p_{T}), the production of prompt charged particles in the backward region is larger than in the forward. The largest difference occurs at $4.0 < |\eta| < 4.3$ and at $p_{\mathrm{T}} \approx 1.5\,\mathrm{GeV}/c$, and corresponds with the largest difference between $4.0 < |\eta| < 4.3$ and $2.5 < |\eta| < 3.0$. Towards higher p_{T}, the R_{BF} values for the different $|\eta|$ ranges tend to converge and become compatible for $6 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$.

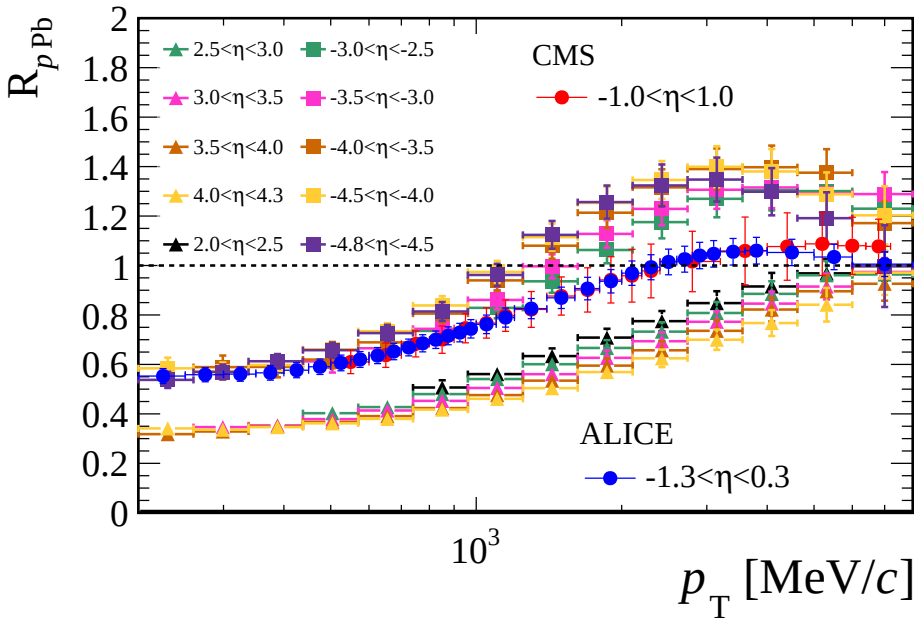


Figure 8.7: Nuclear modification factor for the forward and backward region measured at LHCb and compared with the ALICE [41] and CMS [42,68] measurements in the central region. Error bars include statistical, systematic and luminosity uncertainties added in quadrature.

8.2.1 Comparison with results from other experiments

Fig. 8.7 shows the nuclear modification factor measured by this work and the results in the central region from the ALICE [41] ($-1.3 < \eta < 0.3$) and CMS [42,68] ($-1 < \eta < 1$) collaborations at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. The ALICE and CMS results are in agreement, and they serve as a transition between the different behaviour of the forward and backward $R_{p\mathrm{Pb}}$, although the existence of a considerable gap between the different acceptance intervals must be noted. The gap is of two units of pseudorapidity between CMS and the first LHCb interval in the forward region and also two units between ALICE and the first LHCb interval in the backward region. The Cronin-like enhancement, hinted in the ALICE result, becomes much more pronounced in the backward region.

The result from ATLAS [69] is not included in Fig. 8.7 but is in agreement with the ALICE and CMS result. This measurement has a particularity, which is that the centrality of the proton-lead collision is limited to the $0 - 90\%$ interval. Additionally, ATLAS [69] and ALICE [182] collaborations have also measured particle production in centrality bins in pPb. However, the centrality determination in pPb is subject to important biases which affect the $R_{p\mathrm{Pb}}$ measurement, as is discussed extensively in Ref. [182]. Nevertheless, both measurements indicate enhancements around $p_{\mathrm{T}} \gtrsim 1.5\,\mathrm{GeV}/c$ at the

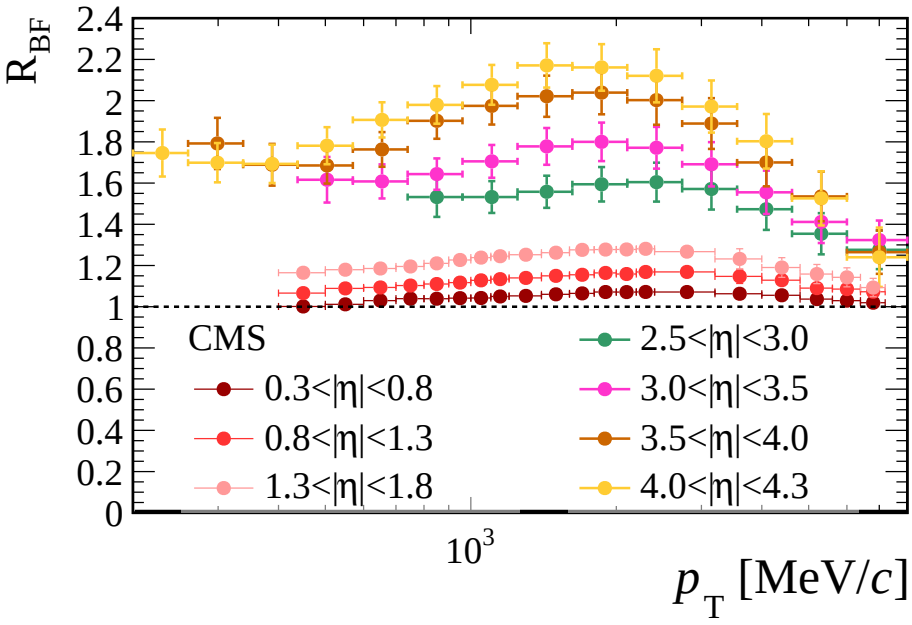


Figure 8.8: Backward-to-forward ratio measured at LHCb compared with the CMS [68] result. For LHCb (non-red points), uncertainties account for statistical, systematic and luminosity sources added in quadrature. The normalisation uncertainty and part of the systematic uncertainty for CMS cancel in their ratio.

highest centrality similarly to what is observed in this work in the backward region. But also, both measurements show strong differences in the $R_{p\mathrm{Pb}}$ for different centrality estimators. Also, ATLAS [69] computes $R_{p\mathrm{Pb}}$ for three η bins, indicating a tendency of stronger enhancement in $p_{\mathrm{T}} \gtrsim 1.5\,\mathrm{GeV}/c$ at more backward η.

Fig. 8.8 shows the backward-to-forward ratio measured in this work compared with the CMS measurement [68] in three η intervals. The gap between the centre of the most forward CMS bin at $1.3 < |\eta| < 1.8$ and the first bin of this work is about 0.5 units of η. Having this in mind, the results are compatible, and the increasing trend of R_{BF} with $|\eta|$ at CMS is continued at LHCb. The R_{BF} trend at high p_{T} is towards no production asymmetry for all η intervals for both experiments.

The $R_{p\mathrm{Pb}}$ can be also compared with measurements performed at RHIC in collisions of protons and deuterium with gold nuclei at the lower energy of $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. The observables for this measurements are $R_{d\mathrm{Au}}$ and $R_{p\mathrm{Au}}$. The gold nucleus has $A = 197$, which is close enough to the lead nuclei with $A = 208$ to expect similar CNM effects. The use of deuterium, a bound state of a proton and a neutron, instead of a proton should also have a minor effect in the nuclear modification factor since the partonic distributions in deuterium are not significantly modified from those in the proton at low x [66, 183, 184]. However, they might be differences if there are important contributions to the charged

particle cross-section from high x partons, where the modification is relevant [66].

The result for $R_{d\mathrm{Au}}$ at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ measured by the BRAHMS collaboration [70] is shown in Fig. 8.9 compared with the measurement of $R_{p\mathrm{Pb}}$ of this thesis. At $\eta = 0$, the BRAHMS measurement shows a similar trend as the LHCb backward data. Then, for forward pseudo-rapidities, the trend of $R_{d\mathrm{Au}}$ resembles the LHCb trend of stronger suppression. Nevertheless, the comparison must be made with caution because of the different collision system, the different energy and the fact that BRAHMS measured exclusively negatively charged hadrons for $\eta = 2.2$ and $\eta = 3.3$. These details have been argued to have significant effects in the nuclear modification [66].

The result for $R_{p\mathrm{Au}}$ at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ measured by the PHENIX collaboration [71] is compared with the $R_{p\mathrm{Pb}}$ from this work in Fig. 8.10. The PHENIX $R_{p\mathrm{Au}}$ results in the forward region, shown in Fig. 8.10a, are in agreement with the $R_{p\mathrm{Pb}}$ of this work for the $2.0 < \eta < 2.5$ and $2.5 < \eta < 3.0$ intervals. The result in the backward region for PHENIX (Fig. 8.10b) shows a behaviour remarkably similar to the most backward $-4.8 < \eta < -4.5$ interval for LHCb. From Fig. 2.7, the overlap region between LHCb and PHENIX backward regions corresponds to the most backward LHCb acceptance, which agrees with the observation of similar trends for the mentioned ranges.

8.2.2 Comparison with phenomenological models

The presented measurements can be compared with the predictions of different phenomenological models. These predictions only cover the $p_{\mathrm{T}} \gtrsim 1.5\,\mathrm{GeV}/c$ region, since they rely on perturbative QCD. Three different calculations are presented:

1) A pQCD calculation [67] based on the nuclear PDF set EPPS16 [75] (for the lead nucleus) and the PDF set CT14 [185] for the proton. The prediction also uses the parton-to-hadron fragmentation function set DSS [26].

2) A prediction based on the CGC effective field theory [89]. The model is only applicable to the saturation region, and thus to forward rapidities (see Sec. 2.4.2). The calculation is leading order, meaning that contributions $\alpha_s = \log(1/x)$ are resumed to all orders. The prediction does not provide an uncertainty estimation.

3) A pQCD calculation within the high-twist factorisation formalism [95, 97]. The calculation predicts a nuclear enhancement of charged hadron production in Pbp collisions in comparison with that in pp collisions in the backward rapidity region, due to the incoherent multiple scattering inside the nucleus before and after the hard scattering. The calculations start from $p_{\mathrm{T}} \approx 2\,\mathrm{GeV}/c$. For $p_{\mathrm{T}} < 2\,\mathrm{GeV}/c$, the pQCD formalism is no longer reliable.

The $R_{p\mathrm{Pb}}$ measurement is compared in Fig. 8.11 with the three listed predictions. The first calculation [67] reproduces the data in the forward region although with large uncertainties. However, it fails to reproduce the $R_{p\mathrm{Pb}}$ enhancement in the backward region for $p_{\mathrm{T}} > 2\,\mathrm{GeV}/c$. The second prediction [89] shows a gradual decrease of $R_{p\mathrm{Pb}}$

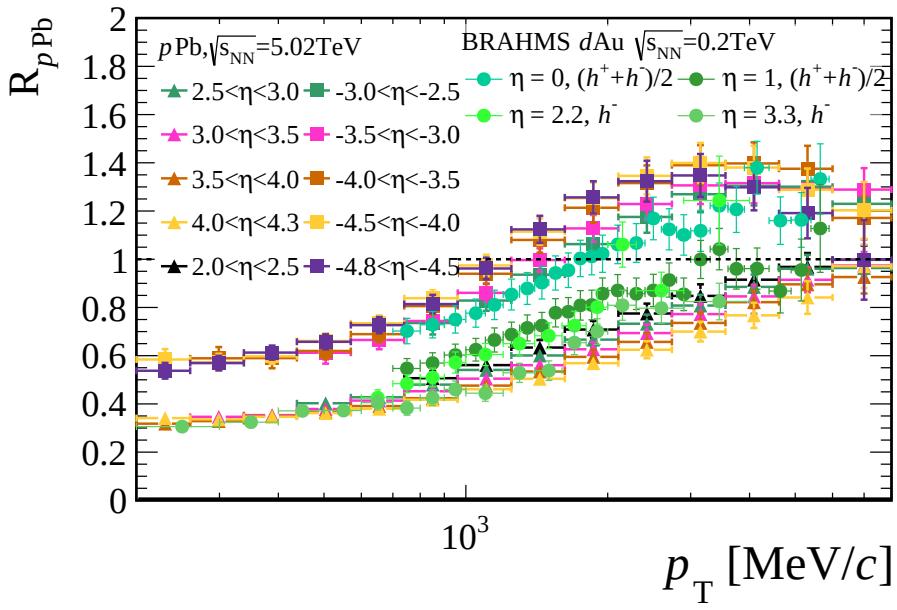
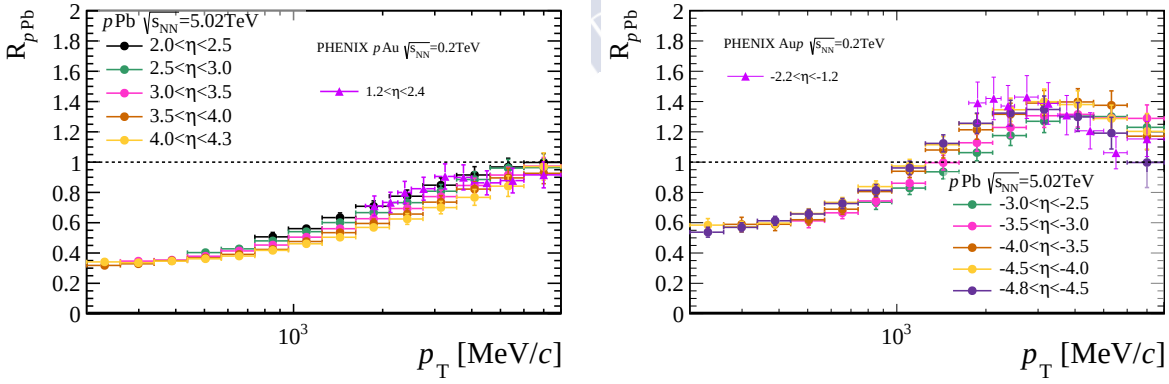


Figure 8.9: Comparison of $R_{p\mathrm{Pb}}$ measured in this work at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ with $R_{d\mathrm{Au}}$ measured by BRAHMS collaboration [70] at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. Error bars account for statistical, systematic and normalisation uncertainties added in quadrature.



(a) (b)

Figure 8.10: Comparison of $R_{p\mathrm{Pb}}$ measured in this work at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ with $R_{p\mathrm{Au}}$ measured by PHENIX collaboration [71] at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$, for forward (Fig. 8.10a) and backward (Fig. 8.10b) regions. Error bars account for statistical, systematic and normalisation uncertainties added in quadrature.

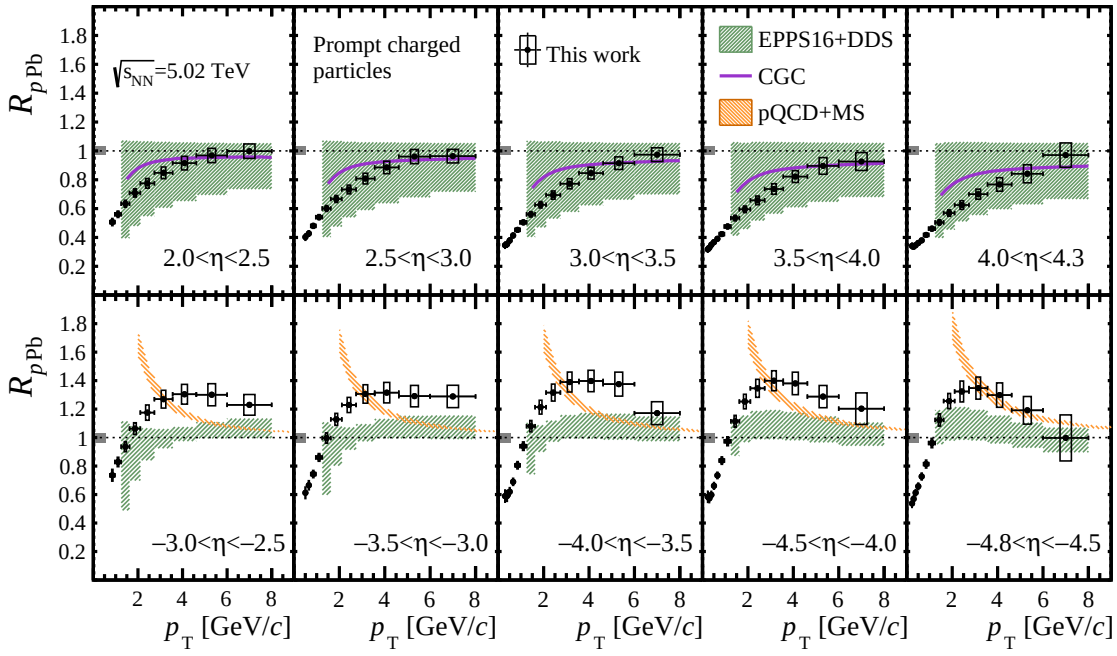


Figure 8.11: Nuclear modification factor for the (top) forward and (bottom) backward regions for different pseudorapidity bins, compared with the predictions mentioned in the text. Vertical error bars correspond to statistical uncertainties, open boxes to uncorrelated systematic uncertainty and the filled box at $R_{p\mathrm{Pb}} = 1$ to the correlated uncertainty from the luminosity.

with η, as observed in data, although it overestimates $R_{p\mathrm{Pb}}$ in the lower p_{T} intervals. The third calculation [95, 97] shows a p_{T} trend similar to data for $p_{\mathrm{T}} > 3\,\mathrm{GeV}/c$ at the most backward η bin, although it does not reproduce the data for the other intervals in the backward configuration. It is worth mentioning that the same framework reproduces the enhancement seen in pAu collisions in the backward region by PHENIX at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ [71] (see Fig. 8.10b).

The prediction from Ref. [67] is also compared with the backward-to-forward and forward-to-backward ratios shown in Fig. 8.12 and Fig. 8.13, respectively. The measured $R_{p\mathrm{Pb}}$ is within the quoted uncertainty bands, except for the first p_{T} intervals around $p_{\mathrm{T}} \approx 2\,\mathrm{GeV}/c$, where the prediction starts. Here, the measured asymmetry is lower (higher) than the prediction for R_{FB} (R_{BF}).

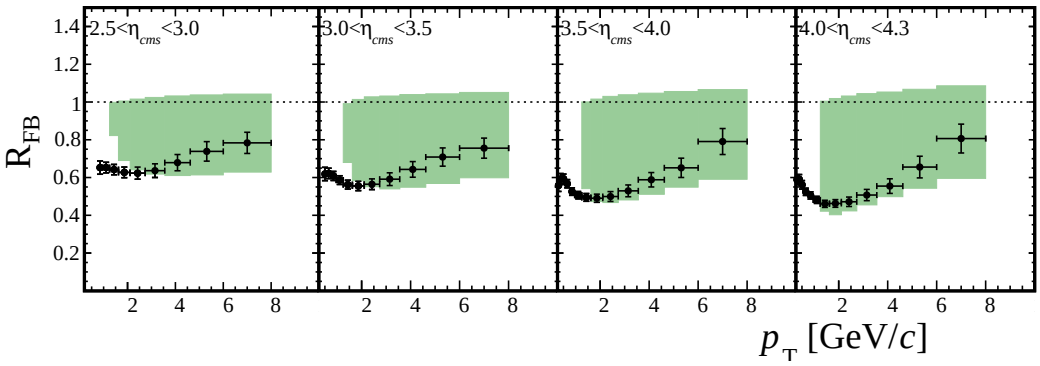


Figure 8.12: Forward-to-backward ratio compared with the calculation from Ref. [67]. Error bars in data account for statistical, systematic and luminosity uncertainties added in quadrature.

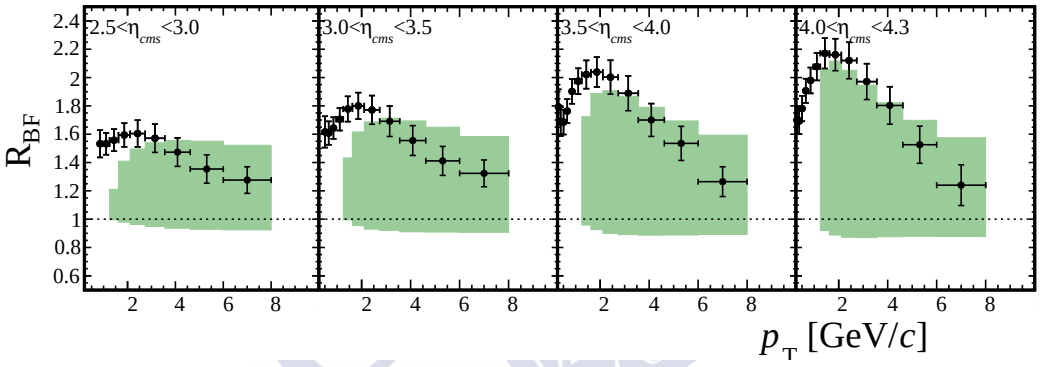


Figure 8.13: Backward-to-forward ratio (bottom) compared with the calculation from Ref. [67]. Error bars in data account for statistical, systematic and luminosity uncertainties added in quadrature.

8.2.3 Dependence with x and Q^2

The understanding of the evolution of $R_{p\mathrm{Pb}}$ with x and Q^2 is a critical point for the study of CNM effects, as explained in Sec. 2.4. However, a direct study presents difficulties because x and Q^2 are partonic quantities that cannot be accessed experimentally.

A simple approach to perform an approximate study is to approximate Q^2 as the transverse mass of the produced particle and to use the expression of Eq. 2.9 to approximate x. With this idea, the experimental variable

$$x_{\mathrm{exp}} \equiv \frac{Q_{\mathrm{exp}}}{\sqrt{s_{\mathrm{NN}}}} e^{-\eta} \tag{8.1}$$

is defined, where $Q^2_{\mathrm{exp}} \equiv m^2 + p^2_{\mathrm{T}}$, and $\sqrt{s_{\mathrm{NN}}}$, p_{T}, η and m refer to the energy of the centre-of-mass system of the collision, the transverse momentum of the produced particles, the

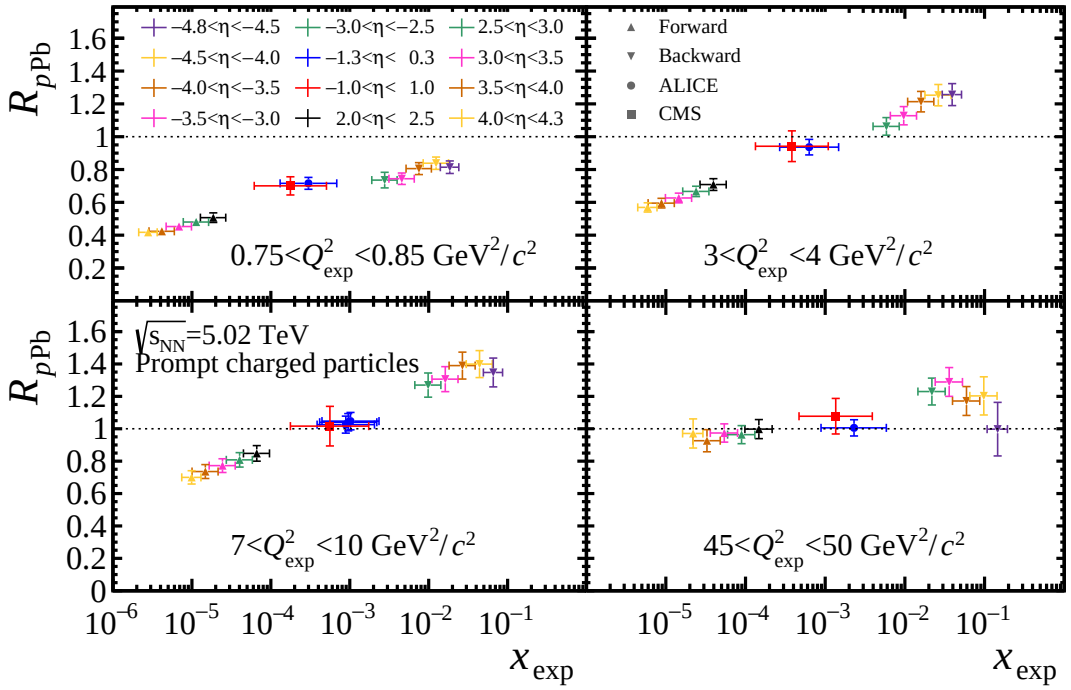


Figure 8.14: Nuclear modification factor as a function of x_{exp}, for the LHCb, ALICE and CMS experiments. The points included in each plot correspond to those whose centre of the p_{T} bin is within the Q^2_{exp} range specified in the plot. Vertical error bars account for the statistical, systematic and the uncertainty in the luminosity (or normalisation) added in quadrature.

pseudorapidity of the produced particles and the mass of the particles. The value of $m = 256\,\mathrm{MeV}/c^2$ is taken, based on an average of the mass of the most abundant particles (pions, kaons and protons) and their relative abundance predicted by the EPOS-LHC generator in the LHCb acceptance.

Every (η, p_{T}) bin can be mapped to a $(x_{\mathrm{exp}}, Q^2_{\mathrm{exp}})$ bin using Eq. 8.1. In Fig. 8.14 the x_{exp} distribution is shown for four different ranges of Q^2_{exp}. Note than the p_{T} binning is different for each experiment, so each plot includes all the $R_{p\mathrm{Pb}}$ p_{T} intervals whose centre is within the mentioned Q^2_{exp} range.

Fig. 8.14 shows the $R_{p\mathrm{Pb}}$ evolution with x_{exp} for four Q^2_{exp} intervals from this study and the results from the ALICE [41] and CMS [42] collaborations. Since the p_{T} binning is different among the three experiments, the Q^2_{exp} ranges are selected to contain at least one p_{T} bin from each experiment. A continuous trend between this measurement in the forward region, the measurements in the central region from ALICE and CMS and the result in the backward region is observed for the four Q^2_{exp} bins. The evolution of $R_{p\mathrm{Pb}}$ with x_{exp} is Q^2_{exp}-dependent. For low Q^2_{exp} $(0.75 < Q^2_{\mathrm{exp}} < 0.85\,\mathrm{GeV}^2/c^2)$ a

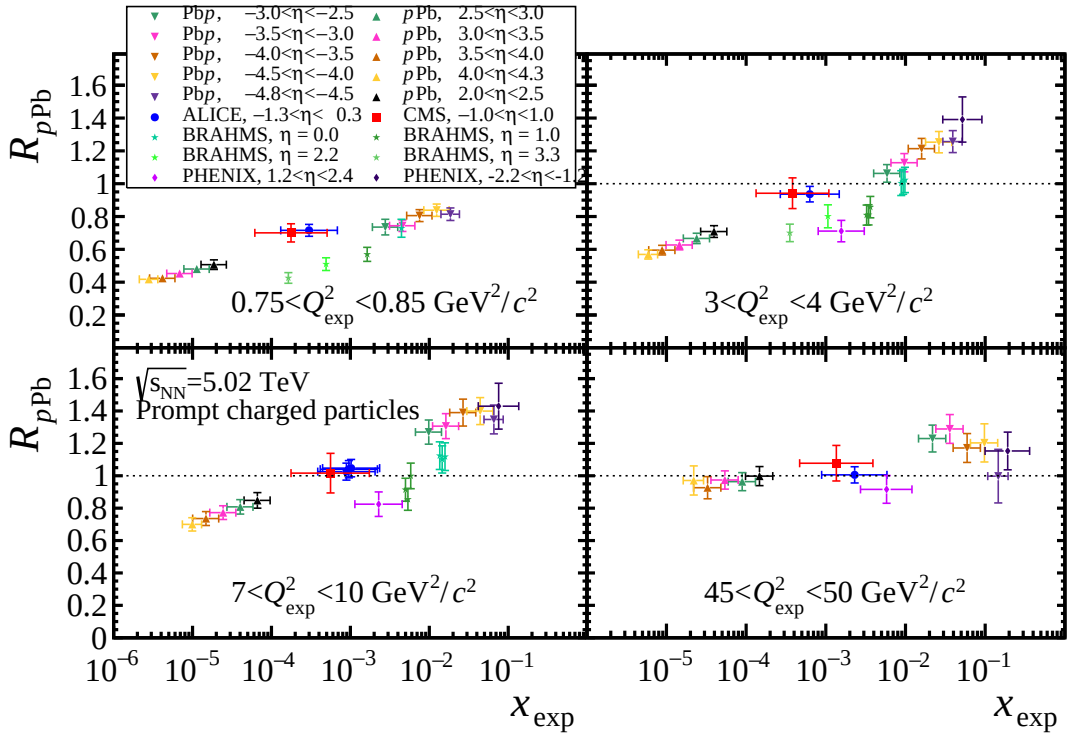


Figure 8.15: Nuclear modification factor as a function of x_{exp}, for LHCb, ALICE, CMS, BRAHMS ($R_{d\mathrm{Au}}$) and PHENIX ($R_{p\mathrm{Au}}$) experiments. Data corresponds to positively and negatively charged particles, except for BRAHMS points for $\eta = 2.2$ and $\eta = 3.3$ which are for negatively charged hadrons only. The points included in each plot correspond to those whose centre of the p_{T} bin is within the Q^2_{exp} range specified in the plot. Vertical error bars account for the statistical, systematic and the uncertainty in the luminosity (or normalisation) added in quadrature.

suppression is observed in every interval. For intermediate Q^2_{exp} ($3 < Q^2_{\mathrm{exp}} < 4\,\mathrm{GeV}^2/c^2$ and $7 < Q^2_{\mathrm{exp}} < 10\,\mathrm{GeV}^2/c^2$) the value of $R_{p\mathrm{Pb}}$ increases strongly from a suppression at low x_{exp} to an enhancement at higher x_{exp}. Finally, for high Q^2_{exp} ($45 < Q^2_{\mathrm{exp}} < 50\,\mathrm{GeV}^2/c^2$) the variation of $R_{p\mathrm{Pb}}$ becomes less x_{exp}-dependent. The start of a decreasing trend of $R_{p\mathrm{Pb}}$ is hinted at $x_{\mathrm{exp}} > 10^{-1}$.

In Fig. 8.15, the previous plot is repeated including $R_{d\mathrm{Au}}$ data from BRAHMS [70] and $R_{p\mathrm{Au}}$ data from PHENIX [71]. The evolution of the nuclear modification factor is less continuous than in Fig. 8.14, but there are several caveats to consider. First, the collision energy is different, which has an impact on the kinematics of the parton being probed, both in the proton (deuterium) and the nucleus. Eq. 2.9 only gives an estimation of the minimum x of probed partons in the nuclei, but as was mentioned in Sec. 2.4, the contributions to the cross-section originate from different x profiles, influencing the cross-section rate [66, 67]. Additionally, the BRAHMS results for $\eta = 2.2$ and $\eta = 3.3$

consider only negatively charged hadrons and deuterium is used instead of proton, which has an impact in the nuclear modification [66]. Having this in mind, the figure manifests the potential of the data from this work to constrain CNM effects in a broad x range. The precision in the backward region is improved with respect to the PHENIX and BRAHMS measurements, adding new constrains to phenomenological models. In the forward region, this measurement covers a region not explored before down to values of $x \sim 10^{-6}$.

9

Conclusions

In this thesis the prompt charged particle production cross-sections in proton-proton and proton-lead collisions at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ have been measured with respect to η and $p_{\rm T}$ in the LHCb experiment. The measurement covers prompt charged particles with $p > 2\,\text{GeV}/c$, $0.2 < p_{\rm T} < 8\,\text{GeV}/c$ and η ranges of $-5.3 < \eta < -2.5$ (Pbp), $1.6 < \eta < 4.3$ (pPb) and $2.0 < \eta < 4.8$ (pp), where η is in the nucleon-nucleon centre-of-mass system. The measured cross-sections have been used to compute the nuclear modification factor $R_{p\text{Pb}}$ for prompt charged particles at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$. The backward-to-forward ratio $R_{\rm BF}$ has been determined as well. The results of this work [178] have been preliminary declared public by the LHCb Collaboration and have been presented at the DIS2021 conference [186]. Shortly, after finishing the internal editorial procedure, the result will be submitted for publication to the Physical Review Letters journal.

Data of proton-lead collisions at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ taken in 2013 were analysed along with data of proton-proton collisions at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ taken in 2015. Prompt charged particle yields were measured from long tracks reconstructed by the LHCb tracking system with a specific selection to improve the signal purity. Then, simulation samples for pPb, Pbp and pp data were validated and used to determine the corrections needed to extract the prompt charged particle spectrum. The main corrections were the reconstruction efficiency, the selection efficiency and the purity. When possible, data-driven techniques were used to calibrate the simulation.

The differential production cross-sections have been determined for the first time in proton-lead collisions in the forward and backward regions at the LHC. The cross-sections for pp collisions at 5.02 TeV have been also determined for the first time, and they are found to be compatible with the recent LHCb preliminary result at 13 TeV [178] and

with the ALICE measurement in the central region [41]. The measured cross-sections provide novel experimental constraints to Monte-Carlo generators of hadronic collisions, which currently present strong discrepancies with data and rely on experimental input to improve their accuracy, as explained in Sec. 2.3. This has a special impact on cosmic ray physics, since current measurements are principally limited by the precision of such generators [5, 6]. In particular, forward data of pp and pPb collisions at LHC energies are specially necessary for this purpose [7].

The nuclear modification factor $R_{p\mathrm{Pb}}$ for charged particles has been determined in the forward and backward regions for the first time at the LHC, representing one of the most precise measurements to date. The total uncertainty, including the normalisation contribution, is below 5% for most of the considered (η, p_{T}) intervals. In the forward region, a strong suppression of the charged particle production is observed, especially for low p_{T} and the most forward η. In the backward region, the measurement has revealed that the interesting mid-p_{T} enhancement of particle production, previously observed at RHIC energies and hinted in ALICE and ATLAS in centrality dependent measurements, is indeed extended to LHC energies in the backward region between $-4.8 < \eta < -2.5$. The $R_{p\mathrm{Pb}}$ shape exhibits a clear pseudorapidity dependence not seen before, which might be the key to give a conclusive explanation to the phenomenon [96]

These data cannot be simultaneously described across the entire measured η range by nPDFs alone. Contrary to what is observed at mid-rapidity [187], the forward data are inconsistent with CGC calculations at the lowest p_{T}. Multiple scattering calculations [95, 97], which successfully reproduce PHENIX results [71], fail to describe the full η range in the backward region. These measurements provide strong constraints on nuclear PDFs at the lowest x ranges accessible in the laboratory, and show that additional, previously unconsidered mechanisms are required to fully describe particle production in nuclear collisions at the LHC.

In the future, the potential of LHCb can be further exploited in a hadron-dependent analysis of the cross-sections, where the π, K and p contributions are separated using the PID capabilities of LHCb explained in Sec. 3.2.3. Previous studies at ALICE [93] and PHENIX [94] have indicated strong differences in the nuclear modification factor of p versus π and K, which could be linked to the physics mechanism behind such enhancement. The differences could be related to the different quark structure, being the proton a baryon (three valence quarks) and pions and kaons mesons (a quark and an antiquark). Apart from its evident theoretical interest, a measurement of the identified spectra will have implications in the inclusive measurement of this thesis, since as mentioned in Chap. 7, the unknown relative particle abundances between π, K and p are at the origin of one of the leading systematic uncertainties.

Future measurements could also investigate differences in the production rate of negative and positive charged hadrons, motivated by the tensions in the comparison with the BRAHMS [70] result with only negative hadrons. Such differences in the nuclear modification have not been seen by any LHC experiment, and would shed light on the underlying mechanism affecting this difference.

To conclude, the original results presented in this thesis improve the current knowledge of QCD in a significant manner. The interest of the measurement cover different areas of research: ranging from heavy-ion phenomenology to cosmic ray physics.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A

Resumo da tese

A.1 Motivación teórica e obxectivos da tese

A Cromodinámica Cuántica (QCD, do inglés "Quantum Chromodynamics") é a teoría fundamental que explica á interacción forte. Esta interacción é a responsable da existencia de hadróns como os protóns ou os neutróns, que se estruturan nunha complexa distribución de partóns (quarks e gluóns). En colisións de hadróns, a taxa de produción de partículas cargadas é un observable fundamental para comprender o funcionamento da QCD.

En colisións de alta enerxía, como as producidas en aceleradores como o Gran Colisor de Hadróns (LHC, polas súas siglas en inglés Large Hadron Collider), as partículas cargadas poden producirse en dous tipos de interaccións entre os partóns: as denominadas febles ("soft") e as duras ("hard"), dependendo de se o intercambio de momento na interacción é maior ou menor con respecto da constante fundamental de QCD, Λ_{QCD}. As interaccións duras poden describirse no marco da QCD perturbativa (pQCD), mentres que o tratamento teórico das interaccións febles é máis complicado e actualmente baséase en consideracións fenomenolóxicas [188, 189]. Incluso ás enerxías acadadas no LHC, as interaccións febles dominan os procesos de produción de partículas cargadas. Por este motivo, os datos experimentais son fundamentais para mellorar os modelos e xeradores de uso común en física de colisores e en física de raios cósmicos [7, 8, 190].

O estudo do réxime de interaccións duras pode facerse a través das partículas cargadas producidas cun alto momento transverso con respecto do eixo no que os hadróns coliden (p_{T}). Este réxime é tratable no marco da QCD perturbativa, e polo tanto proporciona información moi valiosa no eido da física de ións pesados [63]. As diferenzas na produción

de partículas cargadas en colisións protón-chumbo e protón-protón (pp) pódense modelar mediante efectos de materia nuclear fría (CNM, do inglés "cold nuclear matter") [12, 64]. Estes efectos son fundamentais para comprender a física de colisións ión-ión, onde se espera que os efectos CNM aparezan solapados con outros procedentes da formación dun plasma de quark-gluóns (QGP, do inglés "Quark Gluon Plasma"). Ademais, recentemente as indicacións de fenómenos colectivos en sistemas de colisións pequenos como pp ou protón-chumbo poden suxerir a existencia de fenómenos físicos xeralmente non asociados a efectos CNM, senón á formación de QGP [46].

No caso das partículas cargadas, estas modificacións da taxa de produción asócianse xeralmente á efectos de estado inicial, como por exemplo modificacións das funcións de distribución de partóns nun núcleo con respecto ao protón [75–77]. Estas modificacións parametrízanse en funcións nucleares de distribucións de partóns (nPDFs) que se restrinxen mediante axustes aos datos experimentais existentes. Outros efectos nucleares están relacionados con dispersións múltiples dos partóns participantes na colisión [95, 96], as cales poden manifestarse nun fenómeno experimental coñecido como incremento de Cronin [181]. Outras metodoloxías baséanse no fenómeno de saturación de partóns. Tal saturación ocorre cando a densidade de gluóns nos protóns e os núcleos aumenta, o que se corresponde con valores baixos da fracción de momento dos partóns x e con núcleos cun número elevado de nucleóns A [73]. Nesta situación, a dinmica de QCD pode describirse coa teoría de campos efectiva do condensado cristalino de color (CGC, do inglés "colour glass condensate") [85]. As estratexias de nPDFs, CGC e o mecanismo de perda de enerxía [75, 76, 89, 191] poden reproducir a produción de pións na rexión transversa da colisión no LHC ou rexión central [187]. Para acadar valores máis baixos de x, onde aumenta a posibilidade de acadar a saturación, é preciso acceder á rexión de baixo ángulo ou alta pseudorrapidez, definida como:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{A.1}$$

Varias análises realizadas no LHC [41, 42, 69] mediron a produción de partículas cargadas primarias en colisións protón-chumbo á enerxía de centro de masas $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ na rexión central. No Colisor de Ións Pesados Relativistas (RHIC, de "Relativistic Heavy Ion Collider"), realizáronse medidas de colisións dAu e pPb para valores de η superiores pero a enerxías máis baixas ($\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$) [70, 71, 92]. O experimento LHCb, situado no LHC, pode acceder a valores de x máis baixos que calquera outro experimento, xa que cobre a rexión de baixo ángulo ás enerxías propias do LHC.

Nesta tese preséntase unha medida do espectro de partículas cargadas primarias en colisións protón-chumbo e pp a $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ no rango $0.2 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$, cubrindo deste xeito os réximes de produción feble é dura. A medida en colisións protón-chumbo cobre a aceptancia $-5.3 < \eta < -2.5$, cando o feixe de ións chumbo se dirixe cara o espectrómetro LHCb (Pbp, rexión cara atrás), e $2.0 < \eta < 4.3$, cando o feixe de protóns se dirixe cara o espectrómetro LHCb (pPb, rexión cara diante). A medida en colisións pp abrangue $2.0 < \eta < 4.8$. Neste texto, o valor de η está expresado no sistema centro

de masas da colisión nucleón-nucleón, o cal coincide co valor no sistema laboratorio para colisións pp, mentres que se relaciona con este mediante $\eta = \eta_{\rm lab} - 0.465$ nos sistemas pPb e Pbp. Esta diferenza débese ás distintas enerxías dos feixes de protóns e de ións de chumbo no LHC.

O primeiro observable a determinar nesta análise é a sección eficaz diferencial de produción de partículas cargadas, definida como

$$\frac{\mathrm{d}^2\sigma^{\rm ch}(\eta, p_{\rm T})}{\mathrm{d}p_{\rm T}\mathrm{d}\eta} \equiv \frac{1}{\mathcal{L}}\frac{N^{\rm ch}(\eta, p_{\rm T})}{\Delta p_{\rm T}\Delta\eta}. \tag{A.2}$$

Nesta expresión, $N^{\rm ch}$ é o número de partículas cargadas primarias producidas nun intervalo dado de η e $p_{\rm T}$, $\Delta\eta$ e $\Delta p_{\rm T}$, e $\mathcal{L}$ é a luminosidade integrada da mostra de datos correspondente. Neste estudo, unha partícula cargada primaria é calquera hadrón ou leptón cargado con vida media $\tau > 0.3 \times 10^{-10}\,$s producido directamente na colisión ou a partir de desintegracións doutras partículas con vida media máis curta [102]. O factor de modificación nuclear, $R_{p\rm Pb}$, defínese como

$$R_{p\rm Pb}(\eta, p_{\rm T}) \equiv \frac{1}{A}\frac{\mathrm{d}^2\sigma^{\rm ch}_{p\rm Pb}(\eta, p_{\rm T})/\mathrm{d}p_{\rm T}\mathrm{d}\eta}{\mathrm{d}^2\sigma^{\rm ch}_{pp}(\eta, p_{\rm T})/\mathrm{d}p_{\rm T}\mathrm{d}\eta}, \tag{A.3}$$

onde $A = 208$ é o número de nucleóns no ión de chumbo e $\mathrm{d}^2\sigma^{\rm ch}_{p\rm Pb,pp}$ é a sección eficaz diferencial das colisións pPb, Pbp e pp. O último observable a determinar é asimetría entre as rexións cara diante e cara atrás, $R_{\rm BF}$, definida como

$$R_{\rm BF}(\eta, p_{\rm T}) = \frac{\displaystyle\int_{-b}^{-a}\frac{\mathrm{d}^2\sigma^{\rm ch}_{p\rm Pb}(\eta_{\rm cms}, p_{\rm T})}{\mathrm{d}p_{\rm T}\mathrm{d}\eta_{\rm cms}}\mathrm{d}\eta\mathrm{d}p_{\rm T}}{\displaystyle\int_{a}^{b}\frac{\mathrm{d}^2\sigma^{\rm ch}_{p\rm Pb}(\eta_{\rm cms}, p_{\rm T})}{\mathrm{d}p_{\rm T}\mathrm{d}\eta_{\rm cms}}\mathrm{d}\eta\mathrm{d}p_{\rm T}}, \tag{A.4}$$

onde $[a, b]$ é un intervalo determinado de η. O factor $R_{\rm BF}$ equivale ao cociente de $R^{\rm ch}_{p\rm Pb}$ en rexións simétricas de aceptancia en colisións pPb e Pbp, e mide a modificación nuclear relativa entre as dúas aceptancias. Esta cantidade non precisa da medida da sección eficaz en colisións pp.

A.2 O experimento LHCb e mostras de datos

O experimento LHCb é un espectrómetro dun só brazo de baixo ángulo descrito en detalle nas Refs. [4,113]. Os detectores fundamentais para a realización desta análise son un detector de vértice de bandas de silicio (VELO) que rodea a rexión de interacción e permite a medida da posición do punto de colisión, coñecido como vértice primario (VP), un sistema de trazado que proporciona unha medida do momento p das partículas cargadas é dous detectores de anel Cherenkov (RICH) capaces de discriminar entre distintas especies de partículas cargadas.

Nesta análise utilízanse mostras de datos pPb e Pbp adquiridas en febreiro de 2013, cunha luminosidade integrada de $42.7 \pm 1.0\,\mu\text{b}^{-1}$ e $38.7 \pm 1.0\,\mu\text{b}^{-1}$, respectivamente, onde as incertezas non están correlacionadas entre as mostras. Os eventos selecciónanse mediante un sistema de disparo ("trigger") sen nesgos, que só require polo menos unha traza reconstruída no VELO. Ademais, só se consideran eventos cun único VP reconstruído dentro de tres desviacións estándar da media da distribución de VP da mostra total. Esta rexión denomínase rexión luminosa, e correspóndese coa rexión onde se cruzan os feixes.

Por outra banda, a mostra de datos pp foi adquirida en 2015 e consta dunha luminosidade integrada de $3.49 \pm 0.07\,\text{nb}^{-1}$. Os datos selecciónanse cun sistema de disparo sen nesgo que selecciona os cruces de paquetes do feixe que encabezan un tren. Esta estratexia tamén evita a contaminación procedente de paquetes veciños.

Alén da mostra de datos, a análise utiliza mostras simuladas para modelar a eficiencia de reconstrución, os efectos da selección e a contribución de trazas fondo. Na simulación, as colisións pPb e Pbp xéranse usando EPOS-LHC [58], mentres que as colisións pp xéranse con Pythia [52] cunha configuración específica de LHCb [57]. As desintegracións de partículas descríbense con EvtGen [140], mentres que a interacción das partículas co detector e a súa resposta impleméntanse usando Geant4 [48, 142], tal e como se describe na Ref. [139].

A.3 Medida do espectro de partículas cargadas primarias

O obxectivo da análise é medir N^{ch} nas mostras de datos descritas anteriormente. En primeiro lugar, é preciso definir os candidatos a partículas cargadas primarias, que se corresponden con trazas tipo "long". Estas trazas contan con información do VELO e as estacións de trazado situadas tralo dipolo. Deste xeito, o p das partículas medidas está limitado a $p > 2\,\text{GeV}/c$. Na mostra de trazas reconstruídas, aparte de partículas cargadas primarias, pódense atopar trazas fondo. Estas clasifícanse nos seguintes tipos:

- Trazas pantasma. Son defectos da reconstrución que non se corresponden cunha partícula cargada verdadeira. Son especialmente importantes en eventos cunha ocupación do detector alta e a alto p_{T}.
- Trazas clonadas. Tamén son defectos da reconstrución, pero especificamente son parellas de trazas orixinadas por unha única partícula cargada.
- Partículas secundarias. Son trazas producidas por partículas cargadas que non cumpren os requisitos para seren consideradas primarias segundo a definición anterior. Orixínanse principalmente en interaccións de partículas co detector e en desintegracións de partículas primarias.

Para minimizar a presenza destes fondos na mostra de candidatos, aplícase unha selección. As trazas pantasma suprímense utilizando unha selección estrita no resultado

		pp	pPb	Pbp
Selección dos eventos	Liña do sistema de disparo	`Hlt1NoBiasLeadingCrossing`	`Hlt1MBMicroBiasVelo`	`Hlt1MBMicroBiasVelo`
	Tipo de cruce de paquetes	paquete-paquete	paquete-paquete	paquete-paquete
	número de VPs	-	1	1
	posición do VP	-	na rexión luminosa	na rexión luminosa
Selección de candidatos	Orixe das trazas	`TES["Rec/Track/Best"]`	`TES["Rec/Track/Best"]`	`TES["Rec/Track/Best"]`
	Tipo de traza	Long	Long	Long
	η	$2 < \eta < 4.8$	$2 < \eta < 4.8$	$2 < \eta < 4.8$
	p	$p > 2\,\text{GeV}/c$	$p > 2\,\text{GeV}/c$	$p > 2\,\text{GeV}/c$
	p_T	$0.200 < p_\text{T} < 8\,\text{GeV}/c$	$0.200 < p_\text{T} < 8\,\text{GeV}/c$	$0.200 < p_\text{T} < 8\,\text{GeV}/c$
	`pseudoIP`	0.368 mm	0.348 mm	0.348 mm
	`GhostP`	0.078	0.103	0.109
	VELO compartida	`False`	`False`	`False`

Table A.1: Selección considerada na análise, tanto nos eventos (recollida en Sec. A.2) como nos candidatos.

dun algoritmo baseado nunha rede neural, denominado probabilidade pantasma ("ghost probability", `GhostP`) [160]. Ademais, cando dous ou máis candidatos do mesmo evento comparten un segmento reconstruído no VELO, só se considera como candidato a aquel que ten unha mellor calidade do axuste da traza. A traza eliminada denomínase *VELO compartida*. As trazas clonadas son desprezables nas mostras analizadas ($< 0.5\%$). A contaminación de partículas secundarias elimínase utilizando unha variable denominada pseudo parámetro de impacto (`pseudoIP`), que se constrúe como a distancia transversa entre o punto estimado de orixe da traza e a media gaussiana da distribución de VP na mostra total. A vantaxe desta variable é que evita a introdución de nesgos na mostra pp derivados de requirir a reconstrución dun VP. Este criterio é fundamentalmente efectivo para eliminar hadróns producidos en desintegracións de mesóns K^0_S e barións Λ e en interaccións de hadróns co material do detector. Na táboa Táboa A.1 móstrase un resumo dos criterios de selección dos eventos e dos candidatos.

O valor de N^ch pódese obter a partir do número de candidatos $N_\text{candidates}$ mediante a expresión

$$N^\text{ch} = N_\text{candidates} \frac{P}{\varepsilon_\text{reco}\varepsilon_\text{sel}(1/\varepsilon_\text{TM})} C_\text{boost} = \frac{N_\text{candidates}}{\varepsilon_\text{total}}, \tag{A.5}$$

onde ε_reco é a eficiencia de reconstrución, ε_sel é a eficiencia de selección, ε_TM é a eficiencia de combinación traza-partícula na simulación ("truth-matching"), P é a pureza e C_boost é a corrección debido á translación ao sistema centro de masas. Estas correccións estímanse utilizando a simulación e xeralmente corríxense utilizando mostras de calibración extraídas dos datos.

A eficiencia de reconstrución corrixe por trazas non reconstruídas debido a ineficiencias do detector e efectos da súa aceptancia. A eficiencia de trazado obtida coa simulación corríxese utilizando trazas de muóns procedentes da desintegración $J/\psi \to \mu^+\,\mu\text{m}$ que se empregan como referencia para extraer a eficiencia en mostras de datos e simuladas [132] no rango $5 < p < 200\,\text{GeV}/c$. A corrección aplícase en distintos rangos de (η_{lab}, p). Ademáis, como ε_reco depende do nivel de ocupación do detector en cada evento, introdúcense pesos nas mostras simuladas para reproducir a ocupación observada nos

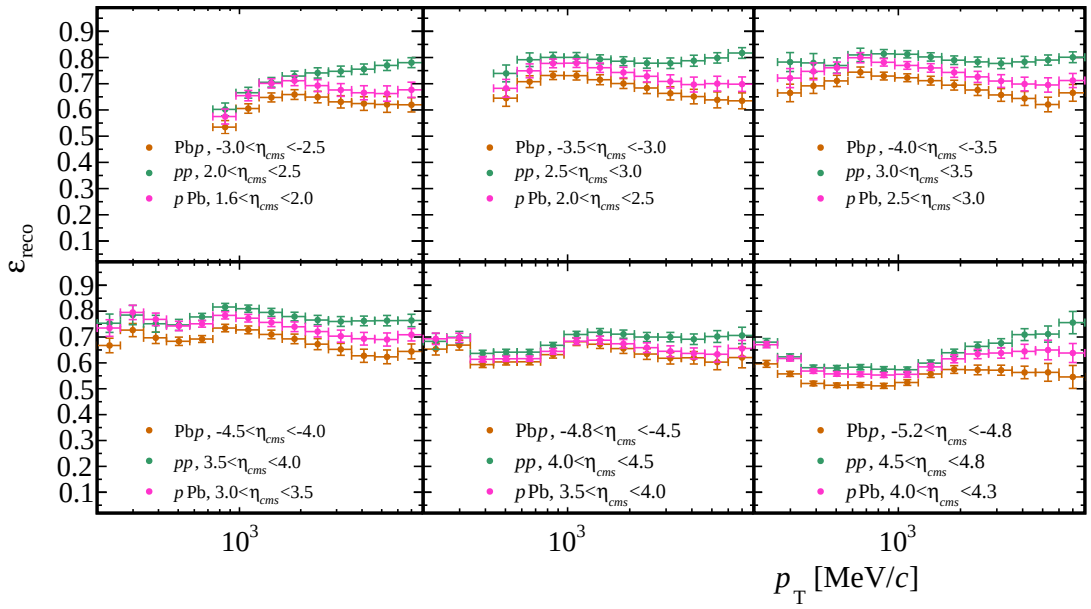


Figure A.1: Eficiencia de reconstrución $\varepsilon_{\text{reco}}$ para colisións Pbp, pPb e pp.

datos adquiridos. Outro punto a considerar é que a eficiencia de reconstrución depende do tipo de partícula. As cantidades relativas de partículas determinadas coa simulación valídanse con datos dos experimentos ALICE [93, 163, 168, 169] e LHCb [166]. A configuración de Pythia de LHCb para pp non reproduce correctamente a cantidade relativa de kaóns e hiperóns primarios na rexión de alto p_{T} nos datos. Polo tanto, unha mostra dedicada de colisións pp, producida con EPOS-LHC [58], utilízase para parametrizar a composición de partículas en pp. As cantidades relativas de EPOS-LHC concordan cos datos mencionados dentro dun 30%, tanto en colisións pp, pPb como en Pbp. Para comprobar a validez da mostra de EPOS-LHC, compáraronse as súas cantidades relativas coas obtidas con Pythia configurado co modelo de hadronización de cordas [167, 192]. A utilización da simulación de Pythia sen corrixir implicaría unha sobreestimación de $\varepsilon_{\text{reco}}$ de ata un 7% a alto p_{T}. O resultado de $\varepsilon_{\text{reco}}$ móstrase na Figura A.1.

A eficiencia de selección corrixe pola fracción de partículas cargadas primarias eliminadas da mostra de candidatos pola selección. Esta eficiencia estímase utilizando a simulación e unha mostra de calibración procedente de desintegracións $\phi(1020) \to K^-K^+$. Esta calibración permite extraer a eficiencia en datos e na simulación, obténdose un factor de corrección para estimar ε_{sel} na mostra de candidatos. O resultado de ε_{sel} móstrase na Figura A.2.

Para obter P, é preciso realizar unha medida das contribucións do fondo. Estas contribucións determínanse coa simulación e corríxense con factores de corrección derivados de mostras de datos enriquecidas en fondo. Para o fondo de trazas pantasma, constrúense dúas mostras independentes, unha utilizando trazas cun valor de `GhostP` elevado

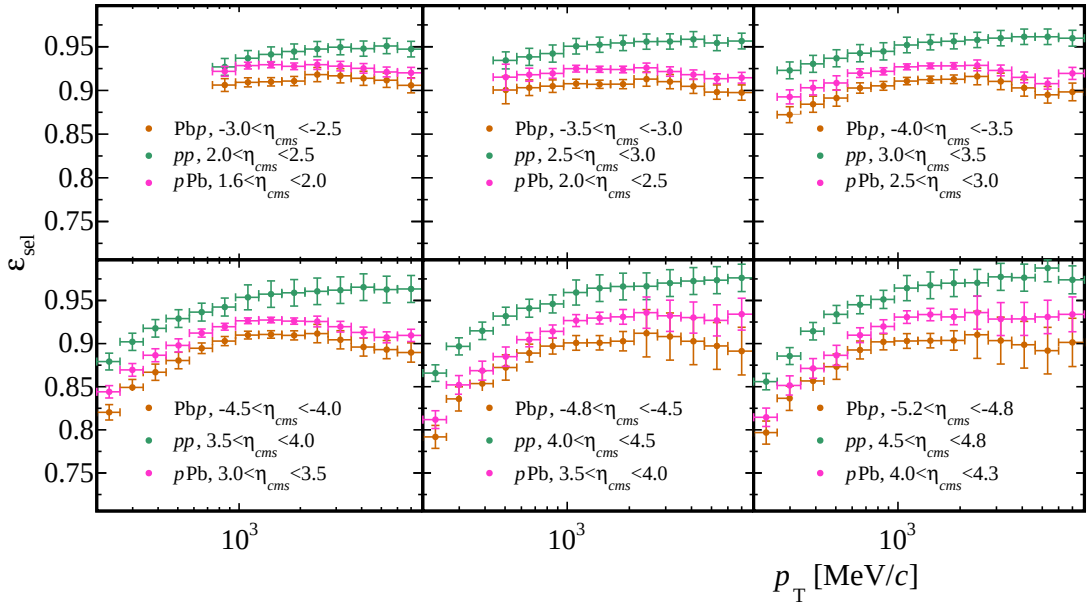


Figure A.2: Eficiencia de selección ε_{sel} para colisións Pbp, pPb e pp.

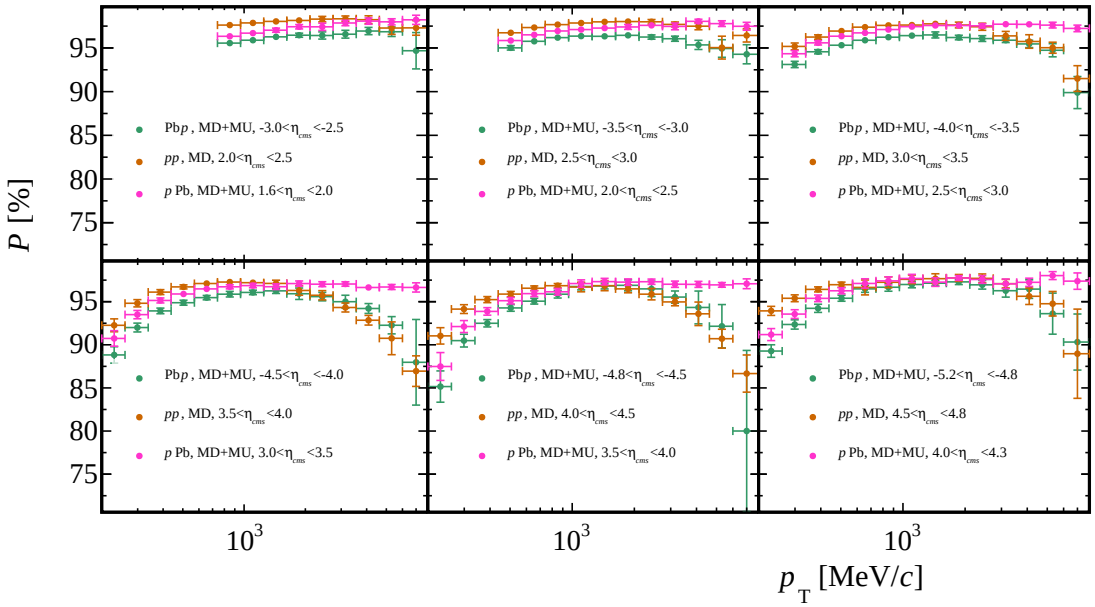


Figure A.3: Pureza P en partículas cargadas primarias da mostra de candidatos en colisións pPb, Pbp e pp.

e outra utilizando trazas tipo VELO compartida. Para o fondo de partículas secundarias, estúdanse as contribucións de hadróns e electróns por separado. Para os hadróns, utilízase unha mostra de trazas con `pseudoIP` elevado. Para os electróns, que son importantes a baixo p_{T} e proceden principalmente de conversións de fotóns co material do detector, a mostra obtense utilizando as variables de identificación de partículas procedentes do sistema RICH. O resultado de P móstrase na Figura A.3.

A eficiencia de combinación traza-partícula na simulación $\varepsilon_{\mathrm{TM}}$ obtense utilizando a mostra de calibración de $\phi(1020) \to K^-K^+$ da simulación, e ten un valor $\sim 1\%$. Os efectos de migración de trazas nos intervalos de (η, p_{T}) debido á resolución do detector son desprezables. Finalmente, o valor de C_{boost} estímase coa simulación, e implica correccións $< 1\%$ a N^{ch} debido á inexactitude da expresión $\eta = \eta_{\mathrm{lab}} - 0.465$ para partículas con masa.

As incertezas que afectan á medida móstranse na Táboa A.2. Para candidatos no rango $5 < p < 200\,\mathrm{GeV}/c$, a eficiencia de trazado leva unha incerteza debido ao tamaño das mostras de calibración e á diferenza entre a interacción dos muóns e os hadróns co material do detector. Para os candidatos fóra deste rango, asígnase unha incerteza do 5%. Outra incerteza asignada considera a descrición da ocupación do detector, que se estima considerando pesos alternativos. A incerteza debida ao descoñecemento da composición relativa dos tipos de partículas determínase mediante a variación das abundancias relativas de partículas obtidas da simulación nun 30% de xeito independente. A incerteza na eficiencia de selección provén principalmente do tamaño das mostras de calibración. Para P, a incerteza sistemática estímase da cantidade de fondo nas mostras enriquecidas e das diferenzas na discrepancia datos-simulación obtida con mostras enriquecidas independentes. Esta incerteza ten unha dependencia con η e p_{T} importante. Mentres que é desprezable en rexións con pouco fondo, domina en rangos de (η, p_{T}) con contribucións importantes ao fondo, como ocorre a alto p_{T} para as trazas pantasma en colisións Pbp. Finalmente, a incerteza asignada a $\varepsilon_{\mathrm{TM}}$ orixínase no tamaño limitado da mostra de calibración.

Orixe da incerteza	pPb [%]	Pbp [%]	pp [%]
Eficiencia de trazado	1.5-5.0	1.5-5.0	1.6-5.3
Ocupación	0.0-2.8	0.6-2.9	0.1-1.6
Composición de partículas	0.4-4.1	0.4-4.6	0.3-2.4
$\varepsilon_{\mathrm{sel}}$	0.7-2.2	0.7-3.0	1.0-1.7
P	0.1-1.8	0.1-11.7	0.1-5.8
$\varepsilon_{\mathrm{TM}}$	0.0-0.1	0.0-0.1	0.1-0.2
$\mathcal{L}^{\dagger}$	2.3	2.5	2.0
Estatística	0.0-0.6	0.0-1.0	0.0-1.1
Total (en $d^2\sigma/d\eta dp_{\mathrm{T}}$)	3.0-6.7	3.3-14.5	2.8-8.7
Total (en $R_{p\mathrm{Pb}}$)	4.2-9.2	4.4-16.9	
Total (en R_{BF})	4.5-11.8		

Table A.2: Incertezas relativas na medida de $d^2\sigma/d\eta dp_{\mathrm{T}}$, $R_{p\mathrm{Pb}}$ e R_{BF} para colisións Pbp, pPb e pp. Os intervalos indican a incerteza máxima e mínima entre todos os rangos de (η, p_{T}). O símbolo † indica unha incerteza correlacionada entre os distintos rangos de (η, p_{T}).

A.4 Resultados e discusión

A sección eficaz diferencial de produción de partículas cargadas primarias en colisións pp e pPb amósase na Figura A.4. A incerteza correspóndese coas incertezas estatística, sistemática e da luminosidade sumadas en cuadratura. En media, prodúcense 0.1035 ± 0.0029 partículas cargadas (con $0.961 < p_{\mathrm{T}} < 1.249\,\mathrm{GeV}/c$ e $3.0 < \eta < 3.5$) en colisións pp, cando o valor medido se escala coa sección eficaz inelástica nucleón-nucleón que vale $67.6 \pm 0.6\,\mathrm{mb}$ a $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ [177]. En colisións pPb, a produción é dúas ordes de magnitude máis alta, asumindo a mesma sección eficaz inelástica nucleón-nucleón. A sección eficaz para colisións pp a $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ pódese comparar co resultado a $13\,\mathrm{TeV}$ [178]. Ámbolos dous resultados son consistentes, observándose un incremento da sección eficaz a $13\,\mathrm{TeV}$ dun factor 1-3, dependendo do p_{T}.

O resultado para o $R_{p\mathrm{Pb}}$ nos distintos intervalos de (η, p_{T}) amósase na Figura A.5, onde as incertezas estatísticas, sistemáticas e da luminosidade preséntanse por separado. Na rexión cara diante, a medida indica unha supresión da produción de partículas cargadas primarias en colisións pPb con respecto a pp, incrementándose cara valores máis altos de η. A baixo p_{T}, $R_{p\mathrm{Pb}}$ acada valores arredor de 0.3 nos ángulos máis baixos. Na rexión cara atrás, obsérvase un incremento significativo sobre a unidade de $R_{p\mathrm{Pb}}$ para $p_{\mathrm{T}} > 1.5\,\mathrm{GeV}/c$. Este incremento pode interpretarse como un incremento tipo Cronin [181], e alcanza un máximo a diferentes valores de p_{T} dependendo de η, seguido por una tendencia decrecente cara a unidade. O decrecemento é máis pronunciado para os valores máis altos de η. O valor máximo de $R_{p\mathrm{Pb}}$ é ~ 1.3 e depende lixeiramente de η.

Na figura Figura A.5 a medida do $R_{p\mathrm{Pb}}$ está comparada con predicións de modelos fenomenolóxicos que cobren a rexión $p_{\mathrm{T}} \gtrsim 1.5\,\mathrm{GeV}/c$. A predición de Ref. [67] está baseada

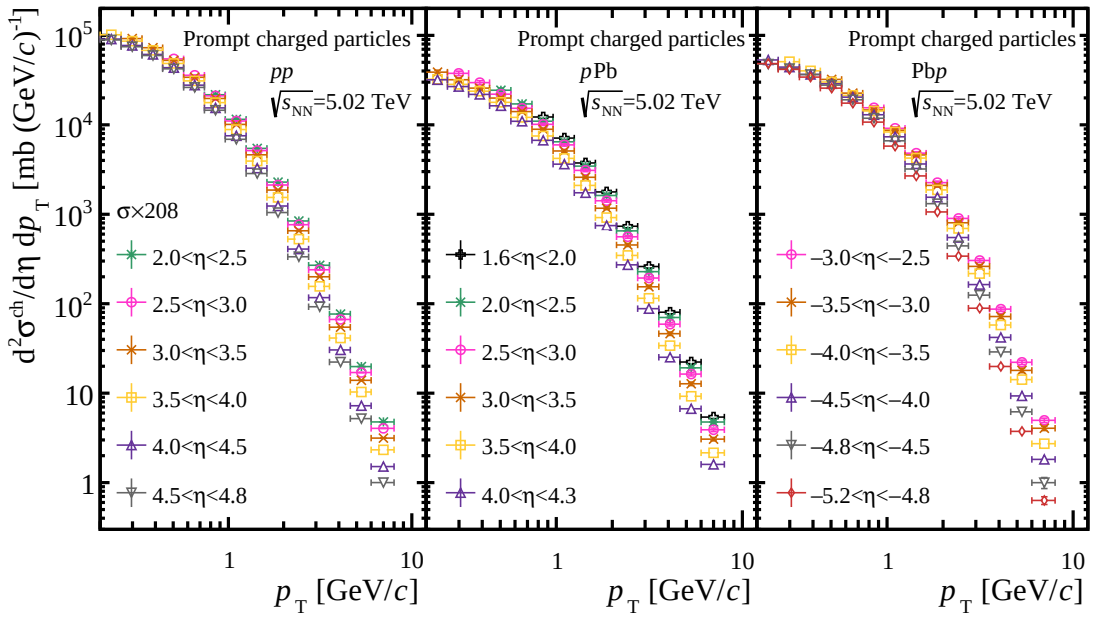


Figure A.4: Sección eficaz diferencial en colisións pp (esquerda), pPb (centro) e Pbp (dereita) a $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. As barras de erro inclúen incertezas estatísticas, sistemáticas e da luminosidade sumadas en cuadratura.

na análise de nPDFs EPPS16 [75] para o núcleo de chumbo e a análise de CT14 [185] para o protón. O cálculo tamén utiliza as funcións de fragmentación partón-hadrón da análise DSS [26]. A predición reproduce os datos medidos na rexión cara diante aínda que con grandes incertezas. Non obstante, fracasa ao reproducir o incremento do $R_{p\mathrm{Pb}}$ na rexión cara atrás para $p_{\mathrm{T}} > 2\,\mathrm{GeV}/c$.

A segunda predición baséase na teoría de campos efectiva CGC [89]. O modelo só se pode aplicar á rexión de saturación e polo tanto na rexión de alta pseudorrapidez. O decrecemento gradual de $R_{p\mathrm{Pb}}$ con η obsérvase nos datos, aínda que a predición sobreestima $R_{p\mathrm{Pb}}$ a baixo p_{T}. A predición non inclúe una estimación da incerteza.

A terceira predición é un cálculo en pQCD no formalismo de factorización "high-twist" na rexión cara atrás [95,97]. O cálculo amosa un incremento debido a dispersións múltiples inelásticas dentro do núcleo antes e despois da interacción dura, e reproduce o incremento observado en colisións pAu na rexión cara atrás polo experimento PHENIX a $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ [71]. A predición amosa unha tendencia con p_{T} similar aos datos para $p_{\mathrm{T}} > 3\,\mathrm{GeV}/c$ no intervalo de pseudorrapidez máis baixo, pero non reproduce correctamente os datos para o resto dos intervalos na rexión cara atrás.

Na figura Figura A.6 móstrase o resultado de R_{BF}. Para todos os intervalos de (η, p_{T}),

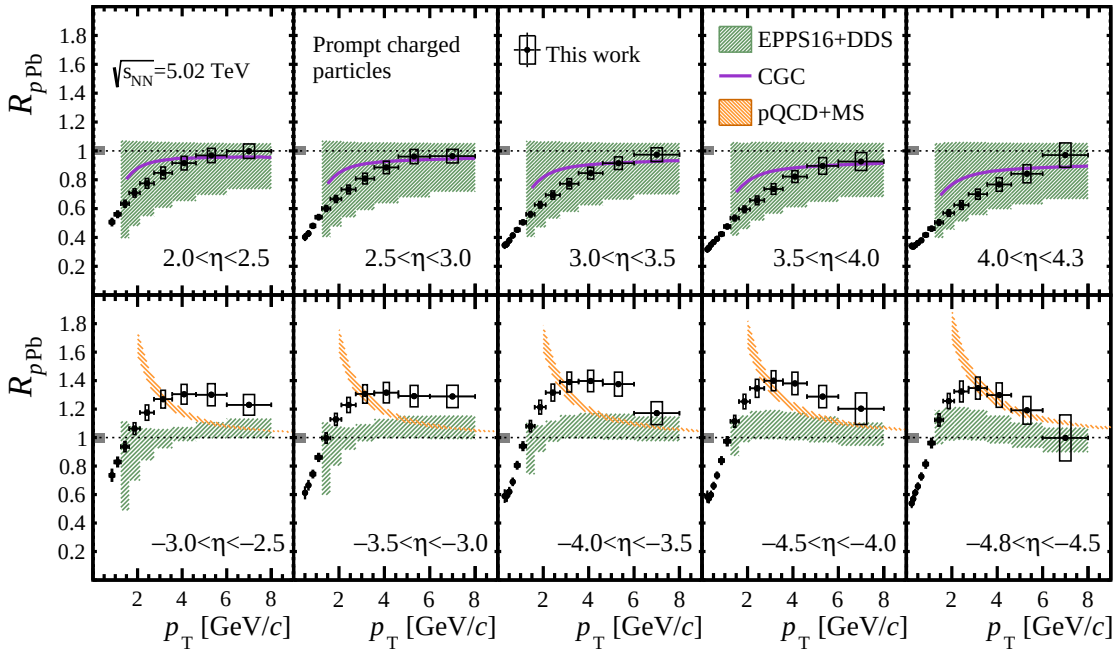


Figure A.5: Factor de modificación nuclear para as rexións (figuras superiores) cara diante e (figuras inferiores) cara atrás para diferentes intervalos de pseudorrapidez, comparado coas predicións mencionadas no texto. As barras de erro verticais (ocultas polo marcador) correspóndese coa incerteza estatística, as caixas non coloreadas coa incerteza sistemática non correlacionada e a caixas coloreadas en $R_{p\mathrm{Pb}} = 1$ coa incerteza correlacionada da luminosidade.

a produción de partículas cargadas primarias na rexión cara atrás é superior a da rexión cara diante. A maior diferenza ocorre para $4.0 < |\eta| < 4.3$ e $p_{\mathrm{T}} \approx 1.5\,\mathrm{GeV}/c$. Cara valores máis altos de p_{T}, R_{BF} para os distintos rangos de $|\eta|$ tende a converxer e é compatible en $6 < p_{\mathrm{T}} < 8\,\mathrm{GeV}/c$. A predición da Ref. [67] tamén se compara coa medida de R_{BF}. A medida atópase dentro das bandas de incerteza da predición, agás para os primeiros intervalos de p_{T} arredor de $p_{\mathrm{T}} \approx 2\,\mathrm{GeV}/c$, onde a predición comeza e subestima o valor de R_{BF}.

Un aspecto fundamental no estudo dos efectos CNM é entender a evolución de $R_{p\mathrm{Pb}}$ con x e o momento transferido entre os partóns Q^2. Non obstante, x e Q^2 son cantidades partónicas e non se poden medir directamente. Polo tanto, considéranse as variables experimentais definidas como

$$Q^2_{\mathrm{exp}} \equiv m^2 + p^2_{\mathrm{T}} \quad \text{and} \quad x_{\mathrm{exp}} \equiv \frac{Q_{\mathrm{exp}}}{\sqrt{s_{\mathrm{NN}}}} e^{-\eta}. \tag{A.6}$$

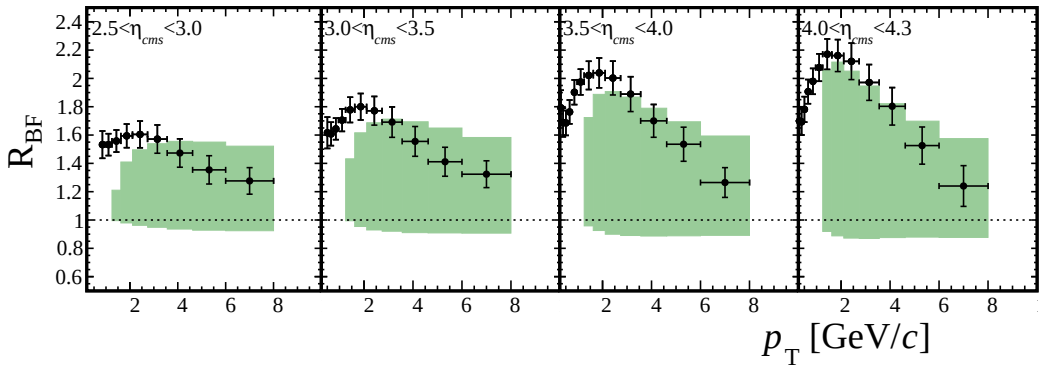


Figure A.6: Valores de R_{BF} medidos e comparados coa predición da Ref. [67]. As barras de erro representan a incerteza estatística, sistemática e da luminosidade sumadas en cuadratura.

Nestas expresións, m é a masa das partículas producidas e tómase como $m = 256\,\mathrm{MeV}/c^2$, a masa promedio das partículas cargadas en colisións pPb determinada con EPOS-LHC. A variable x_{exp} é aproximadamente x para unha dispersión de dous partóns, e Q_{exp} é a masa transversa da partícula producida.

A Figura A.7 mostra a evolución de $R_{p\mathrm{Pb}}$ con x_{exp} para catro intevalos de Q^2_{exp} cos resultados deste estudo e as medidas das colaboracións ALICE [41] e CMS [42]. Como os intervalos de p_{T} son diferentes para os tres experimentos, os rangos de Q^2_{exp} selecciónanse para conter, polo menos, un intervalo de p_{T} de cada experimento. Obsérvase unha tendencia consistente entre a medida na rexión cara diante, as medidas na rexión central de ALICE e CMS e o resultado na rexión cara atrás nos catro intervalos de Q^2_{exp}. Para baixo Q^2_{exp} ($0.75 < Q^2_{\mathrm{exp}} < 0.85\,\mathrm{GeV}^2/c^2$) obsérvase unha supresión para todo x_{exp}. Para Q^2_{exp} intermedio ($3 < Q^2_{\mathrm{exp}} < 4\,\mathrm{GeV}^2/c^2$ e $7 < Q^2_{\mathrm{exp}} < 10\,\mathrm{GeV}^2/c^2$) o valor de $R_{p\mathrm{Pb}}$ aumenta fortemente dunha supresión a baixo x_{exp} cara a un incremento a x_{exp} máis alto. Para alto Q^2_{exp} ($45 < Q^2_{\mathrm{exp}} < 50\,\mathrm{GeV}^2/c^2$) $R_{p\mathrm{Pb}}$ varía menos con x_{exp}. Insinúase o comezo dunha tendencia decrecente de $R_{p\mathrm{Pb}}$ a $x_{\mathrm{exp}} > 10^{-1}$.

Para concluír, a tese presenta a medida das seccións eficaces diferenciais de produción de partículas cargadas primarias con respecto de η e p_{T} en colisións pp, pPb e Pbp a $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. A medida abarca os rangos de $p > 2\,\mathrm{GeV}/c$ e $0.2 < p_{\mathrm{T}} < 8.0\,\mathrm{GeV}/c$ con $2.0 < \eta < 4.8$ en colisións pp, $-5.3 < \eta < -2.5$ en Pbp e $1.5 < \eta < 4.3$ en pPb. Esta é a primeira determinación destas seccións eficaces en colisións protón-chumbo nas rexións cara diante e cara atrás, e a primeira medida en colisións pp a $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. A incerteza total está cerca de 3% para a maioría dos intervalos cinemáticos, tanto en colisións pp como Pbp e pPb. Deste xeito, os datos establecen fortes restricións aos modelos de QCD non perturbativos en colisións hadrónicas a altas enerxías.

O factor de modificación nuclear $R_{p\mathrm{Pb}}$ para partículas cargadas primarias tamén foi determinado e constitúe unha das medidas máis precisas ata o momento. A incerteza

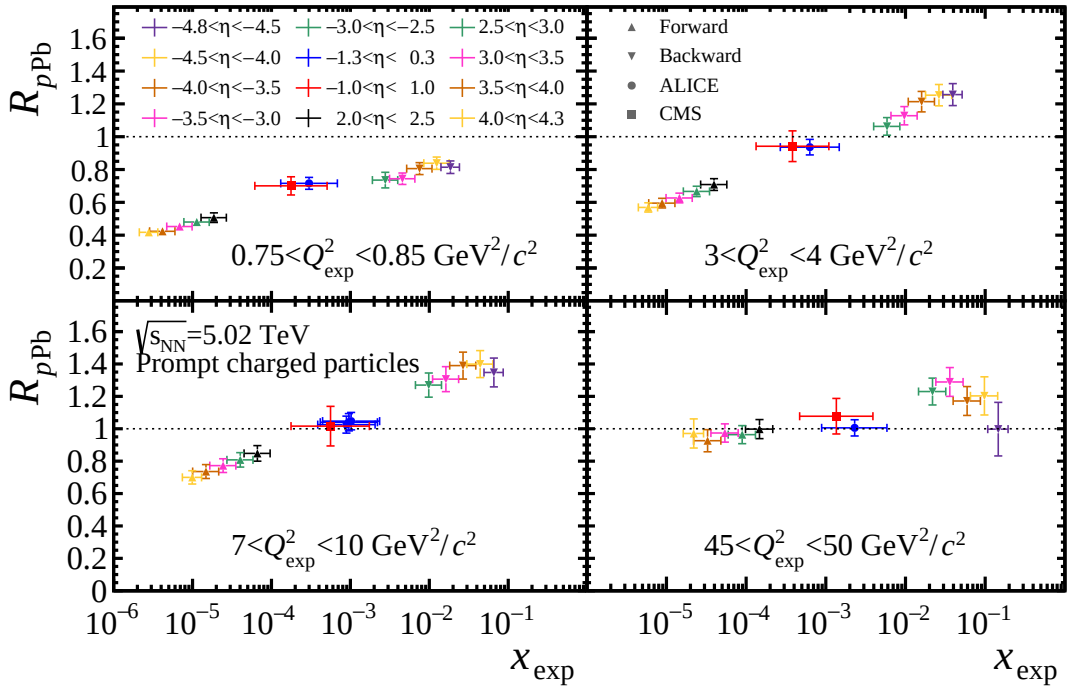


Figure A.7: Factor de modificación nuclear en función de x_{exp}, para os experimentos LHCb, ALICE e CMS. Os puntos incluídos en cada figura correspóndense a aqueles cuxo centro do intervalo de p_{T} está dentro do rango especificado de Q^2_{exp}. As barras de erro consideran a incerteza estatística, sistemática e a da luminosidade (ou normalización) sumadas en cuadratura.

total, incluíndo a contribución da normalización, está por debaixo do 5% para a maioría dos intervalos de (η, p_{T}). Na rexión cara diante, obsérvase unha supresión da produción de partículas cargadas, especialmente para baixo p_{T} e os valores de η máis altos. Na rexión cara atrás, obsérvase un incremento da produción de partículas cargadas con $p_{\mathrm{T}} > 1.5\,\mathrm{GeV}/c$. A forma de $R_{p\mathrm{Pb}}$ indica unha clara dependencia coa pseudorrapidez. Estes datos non se poden describir simultaneamente utilizando nPDFs en todo o rango medido de η. Os datos son inconsistentes coa predición de CGC ao p_{T} máis baixo, ao contrario do que se observa na rexión central [187]. Os cálculos que consideran dispersións múltiples dos partóns non poden describir a rexión cara atrás, mentres que describen os resultados de PHENIX [71]. Esta medida proporciona fortes restricións aos modelos de nPDFs nos rangos de x máis pequenos actualmente accesibles, e implican a necesidade de considerar novos mecanismos para proporcionar unha descrición consistente da produción de partículas en colisións nucleares no LHC.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

B

Additional figures

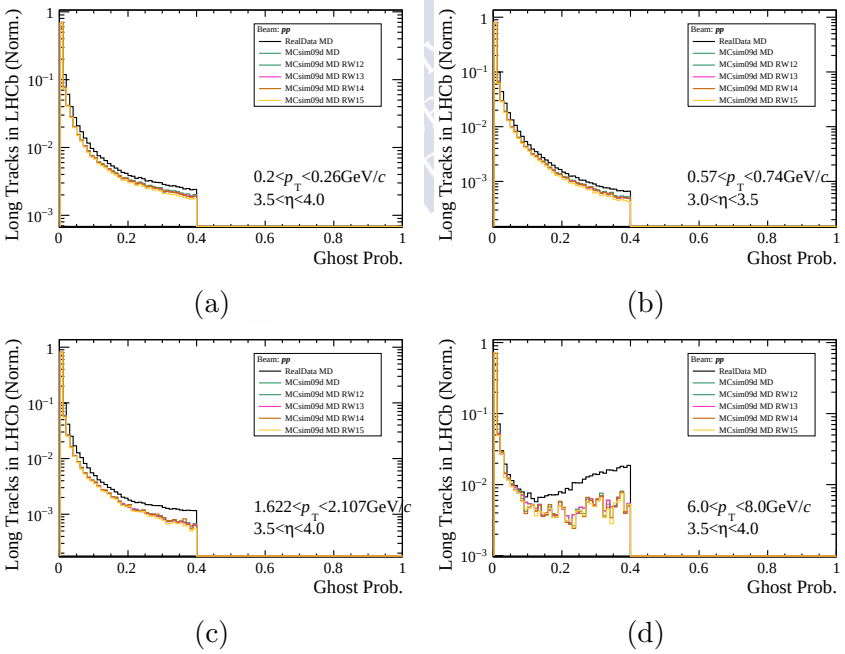


Figure B.1: Distributions of long tracks in the LHCb acceptance with respect to `GhostP` for particular (η, p_{T}) ranges in data and simulation for the pp dataset.

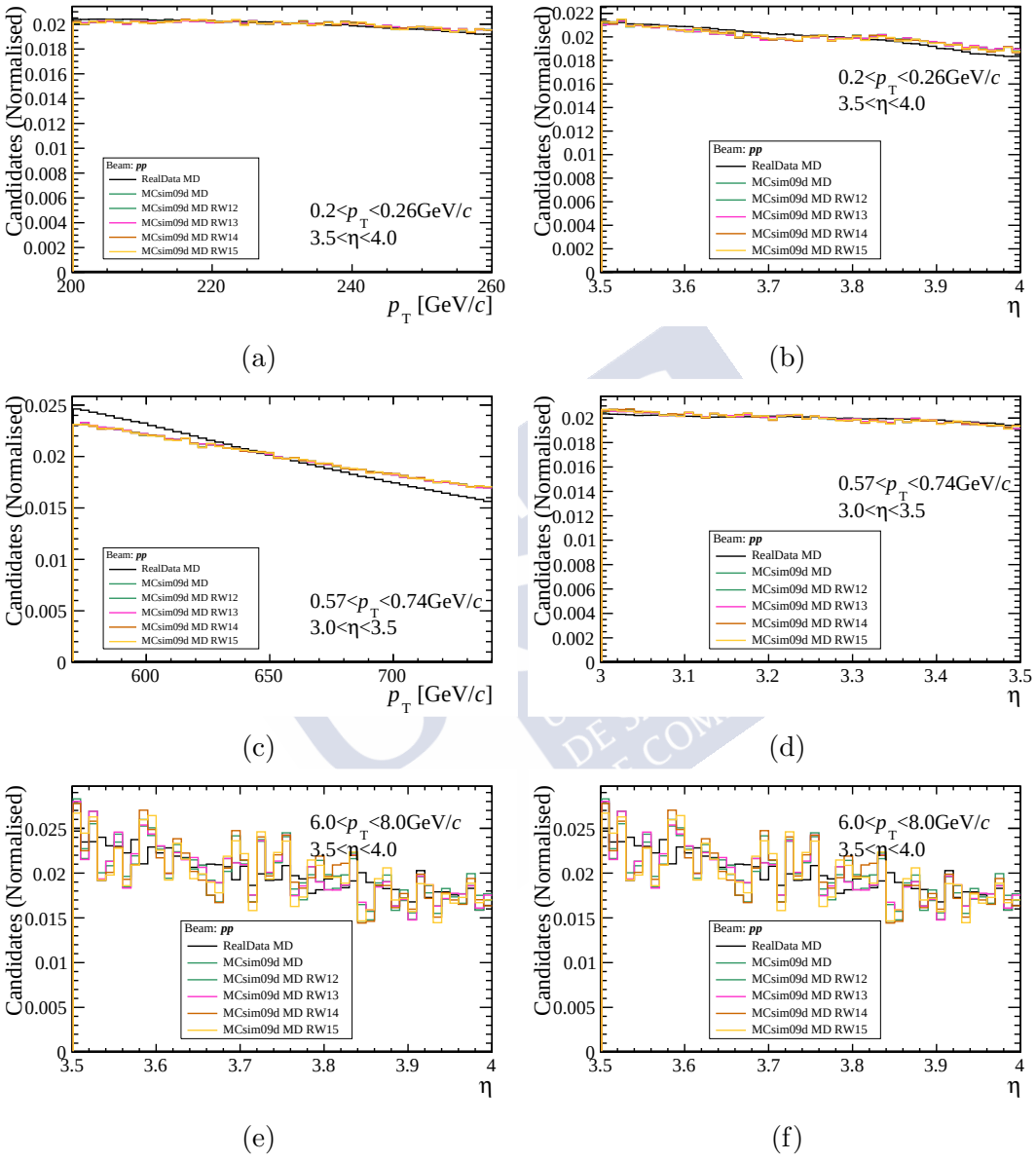


Figure B.2: Distributions of candidates with respect to p_{T} and η for particular (η, p_{T}) ranges in data and simulation for the pp dataset.

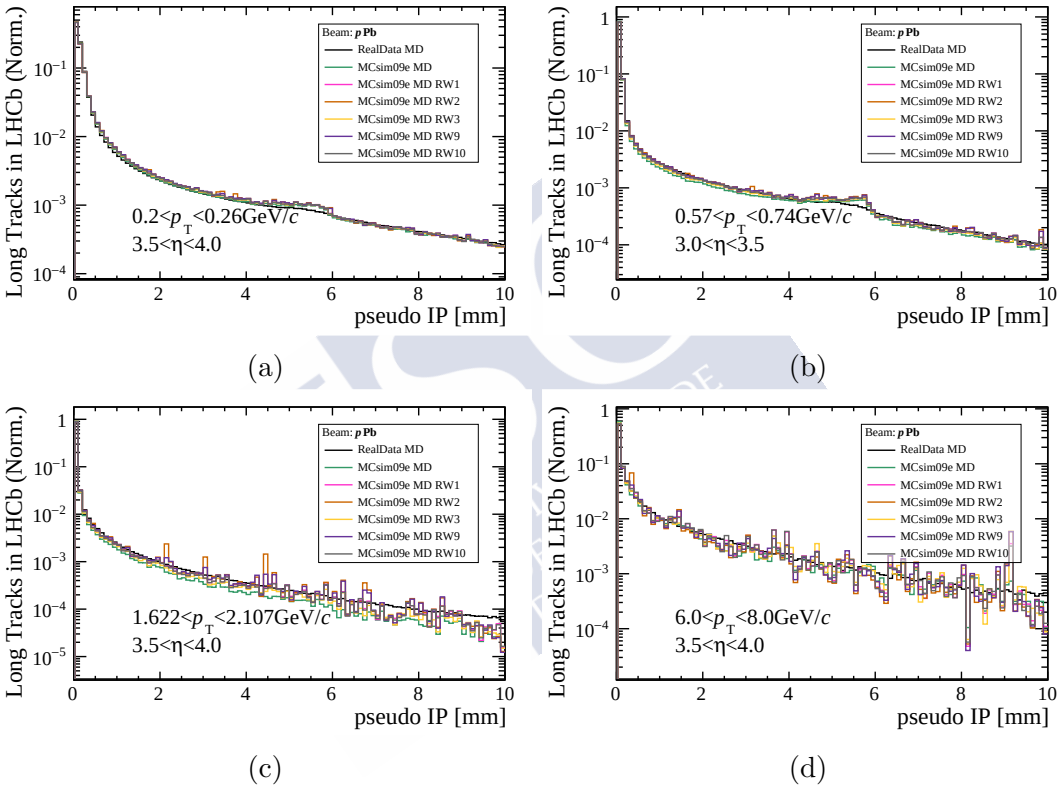


Figure B.3: Distributions of long tracks in the LHCb acceptance with respect to `pseudoIP` for particular (η, p_{T}) ranges in data and simulation for the pPb dataset.

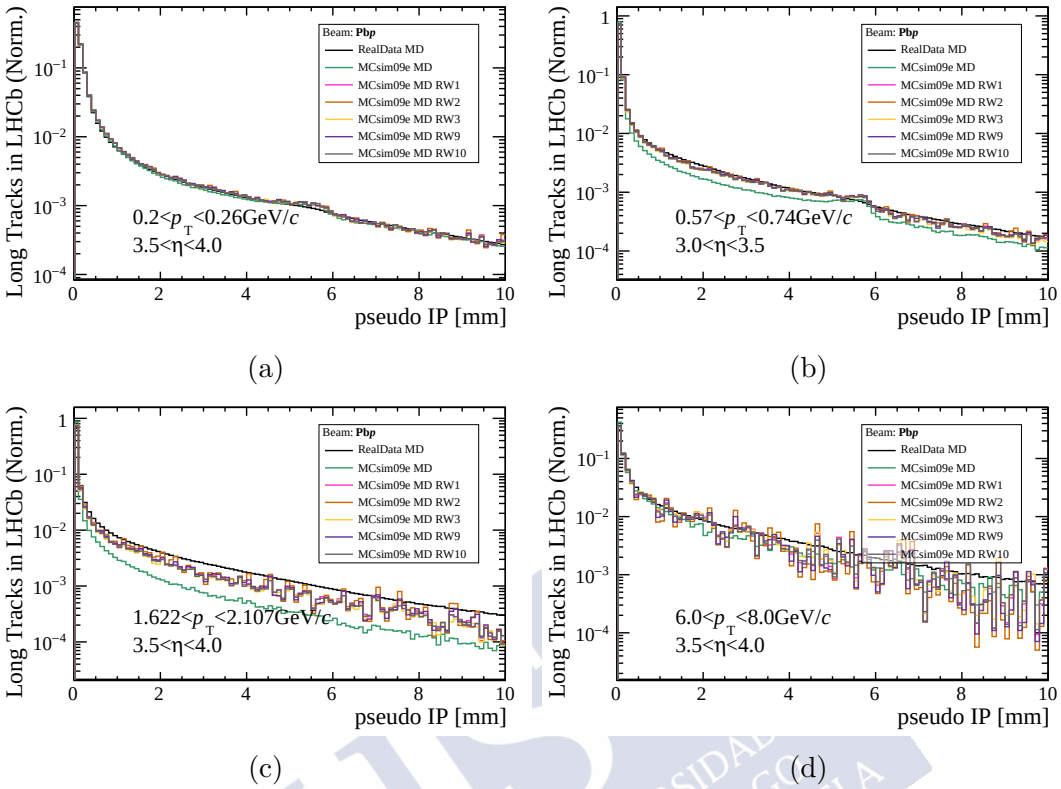


Figure B.4: Distributions of long tracks in the LHCb acceptance with respect to `pseudoIP` for particular (η, p_{T}) ranges in data and simulation for the Pbp dataset.

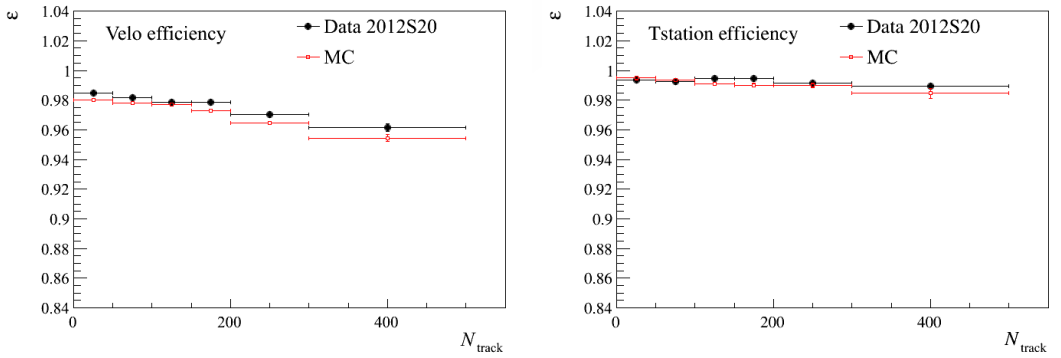


Figure B.5: Tracking efficiency as a function of N_{tracks} in data and simulation for the 2012 pp calibration samples. Left figure (Fig. B.5a) shows the result from the VELO method, while right figure (Fig. B.5b) shows the result from the Tstation method. Figures from Ref. [132].

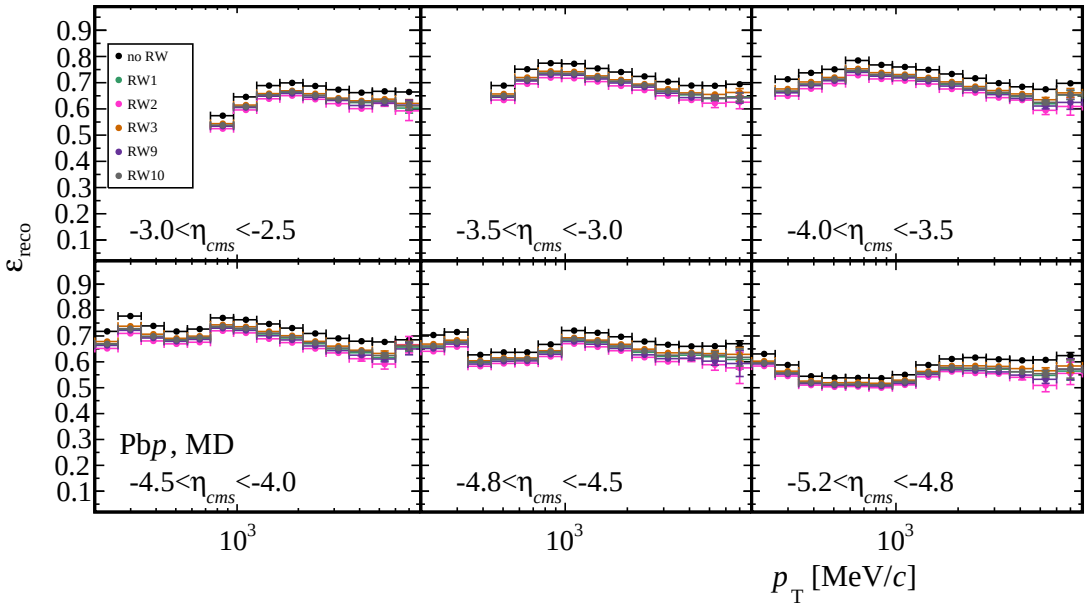


Figure B.6: Variation of the reconstruction efficiency in every bin considering the different weight sets for the Pbp configuration.

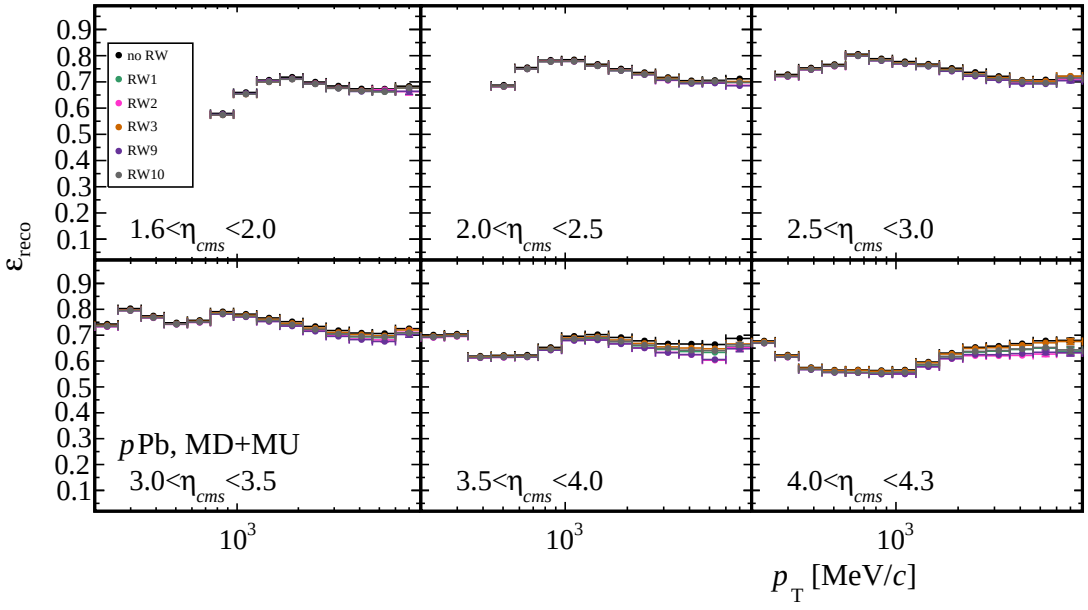


Figure B.7: Variation of the reconstruction efficiency in every bin considering the different weight sets for the pPb configuration.

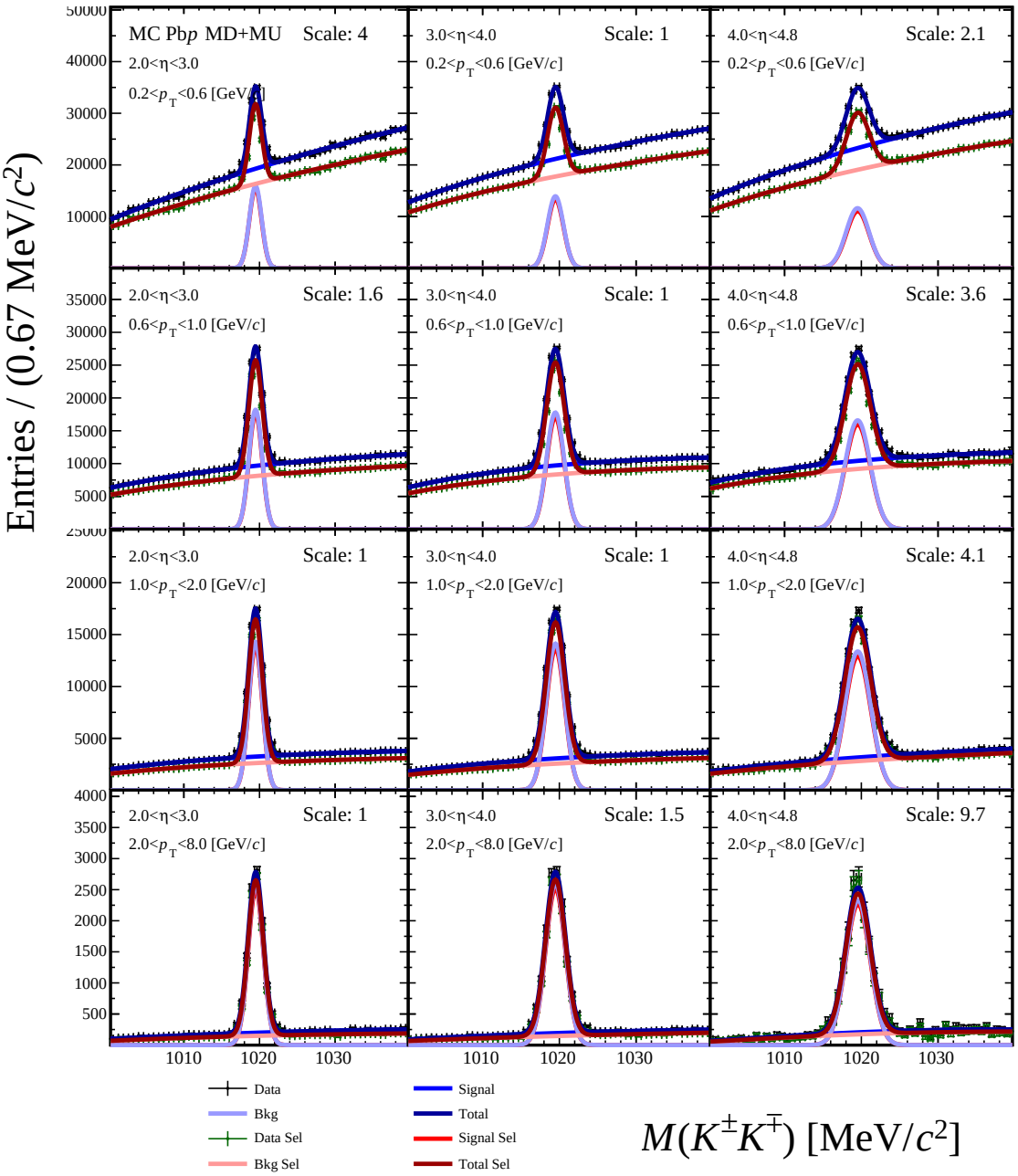


Figure B.8: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ for the simulation sample of the Pbp collisions in the different (η, p_{T}) kinematic bins. The (η, p_{T}) values are indicated in the top left side of each plot. Some plots are re-scaled to make them appear in the centre of the plot, the scale factor is indicated in the top right side.

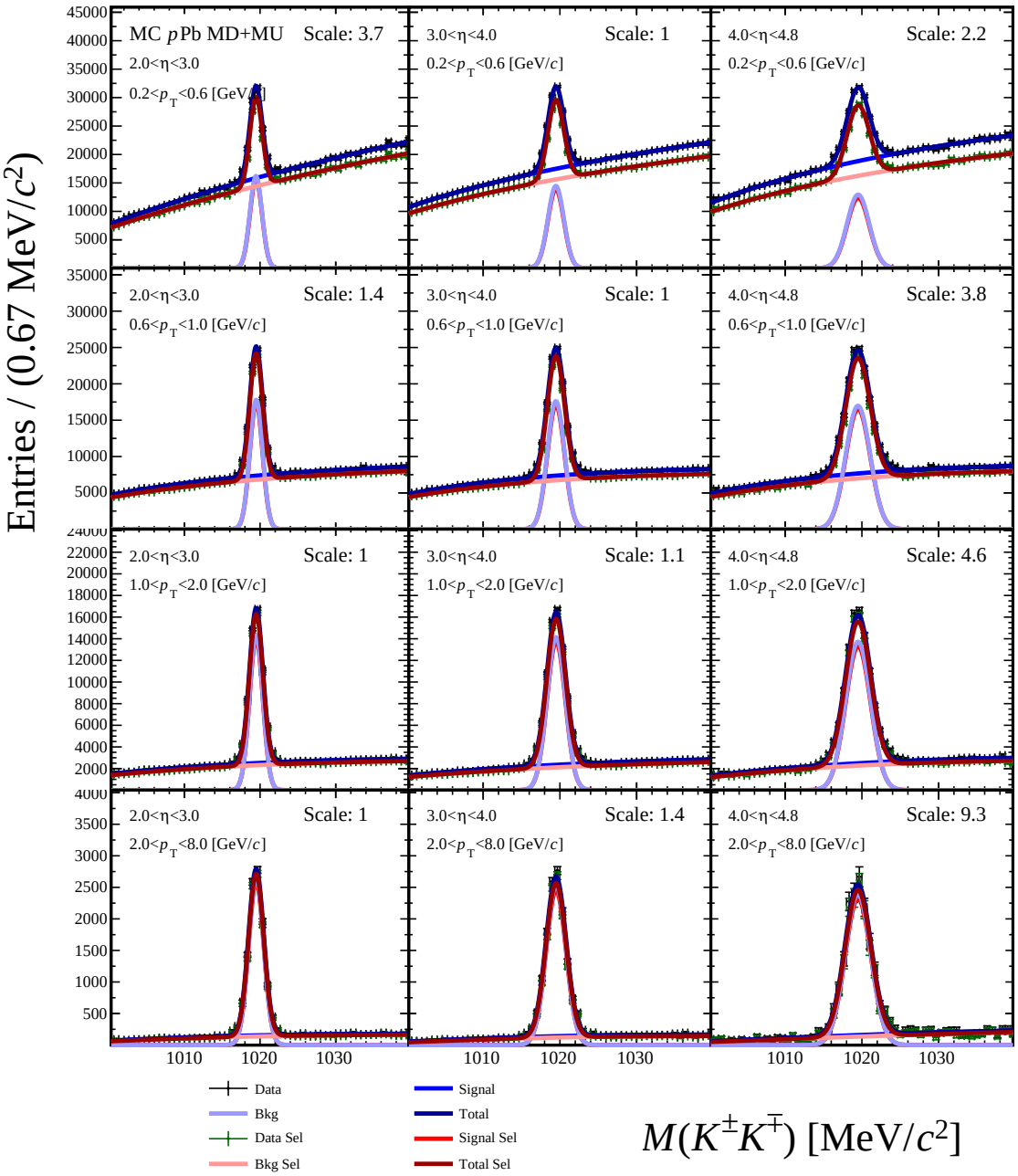


Figure B.9: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ for simulation sample of pPb collisions for the different kinematic bins (η, p_{T}). The (η, p_{T}) values are indicated in the top left side of each plot. Some plots are re-scaled to make them appear in the centre of the plot, the scale factor is indicated in the top right side.

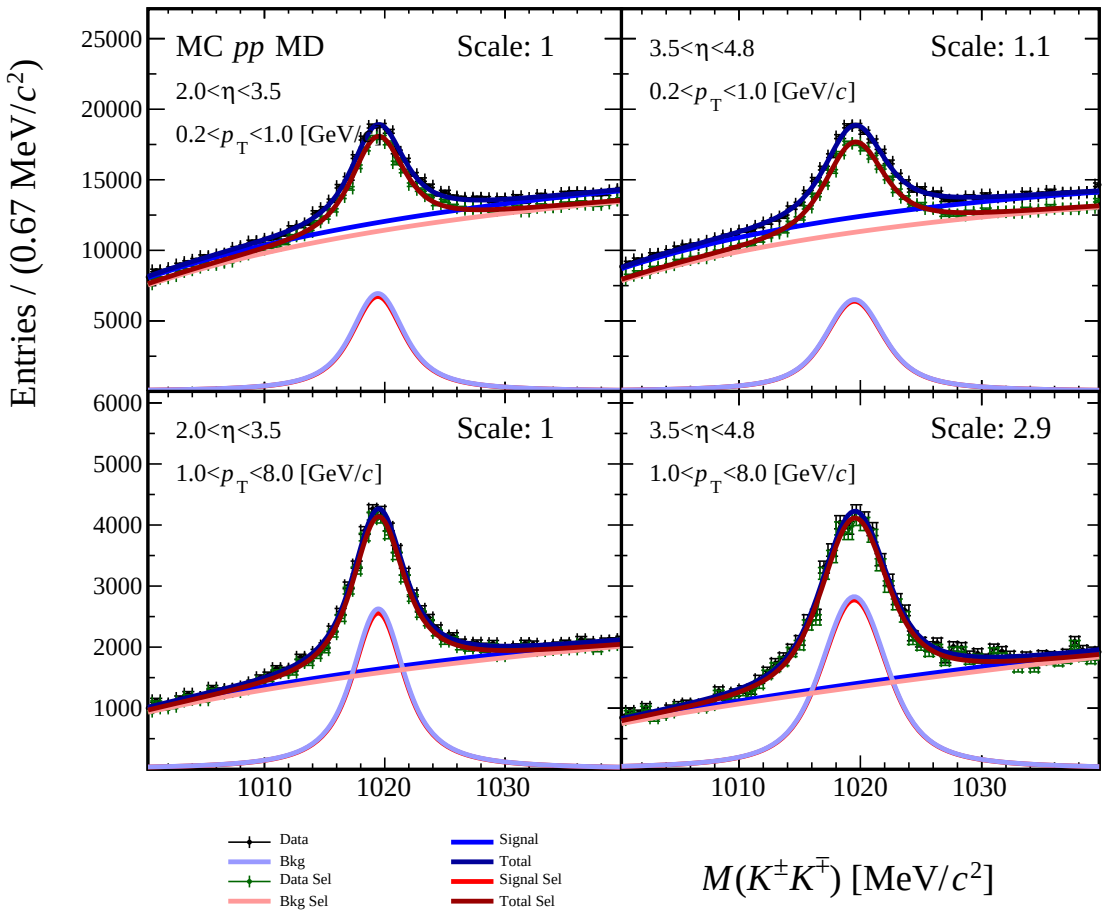


Figure B.10: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ for simulation sample of pp collisions for the different kinematic bins (η, p_{T}). The (η, p_{T}) values are indicated in the top left side of each plot. Some plots are re-scaled to make them appear in the centre of the plot, the scale factor is indicated in the top right side.

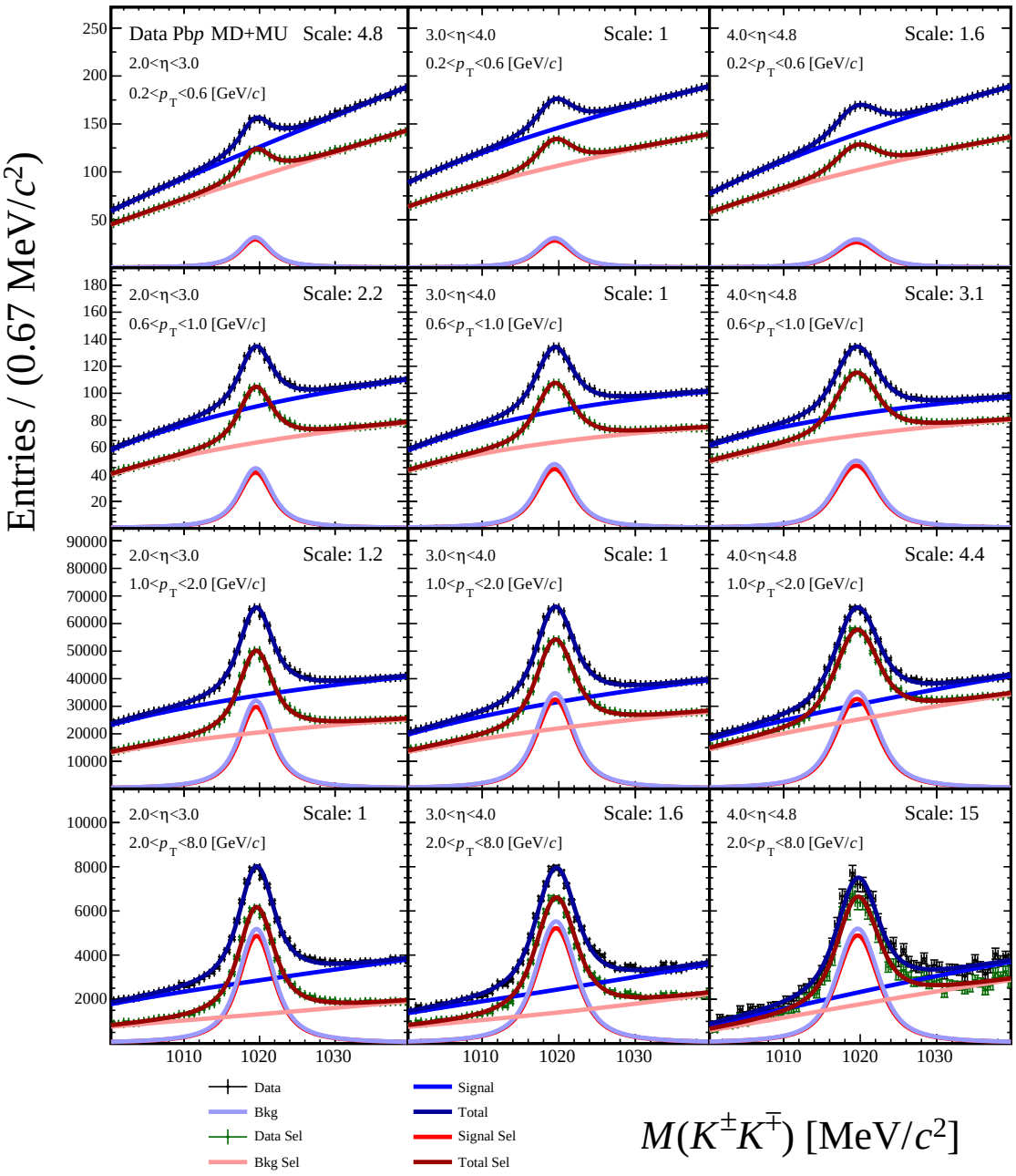


Figure B.11: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ for data sample of Pbp collisions for the different kinematic bins (η, p_{T}). The (η, p_{T}) values are indicated in the top left side of each plot. Some plots are re-scaled to make them appear in the centre of the plot, the scale factor is indicated in the top right side.

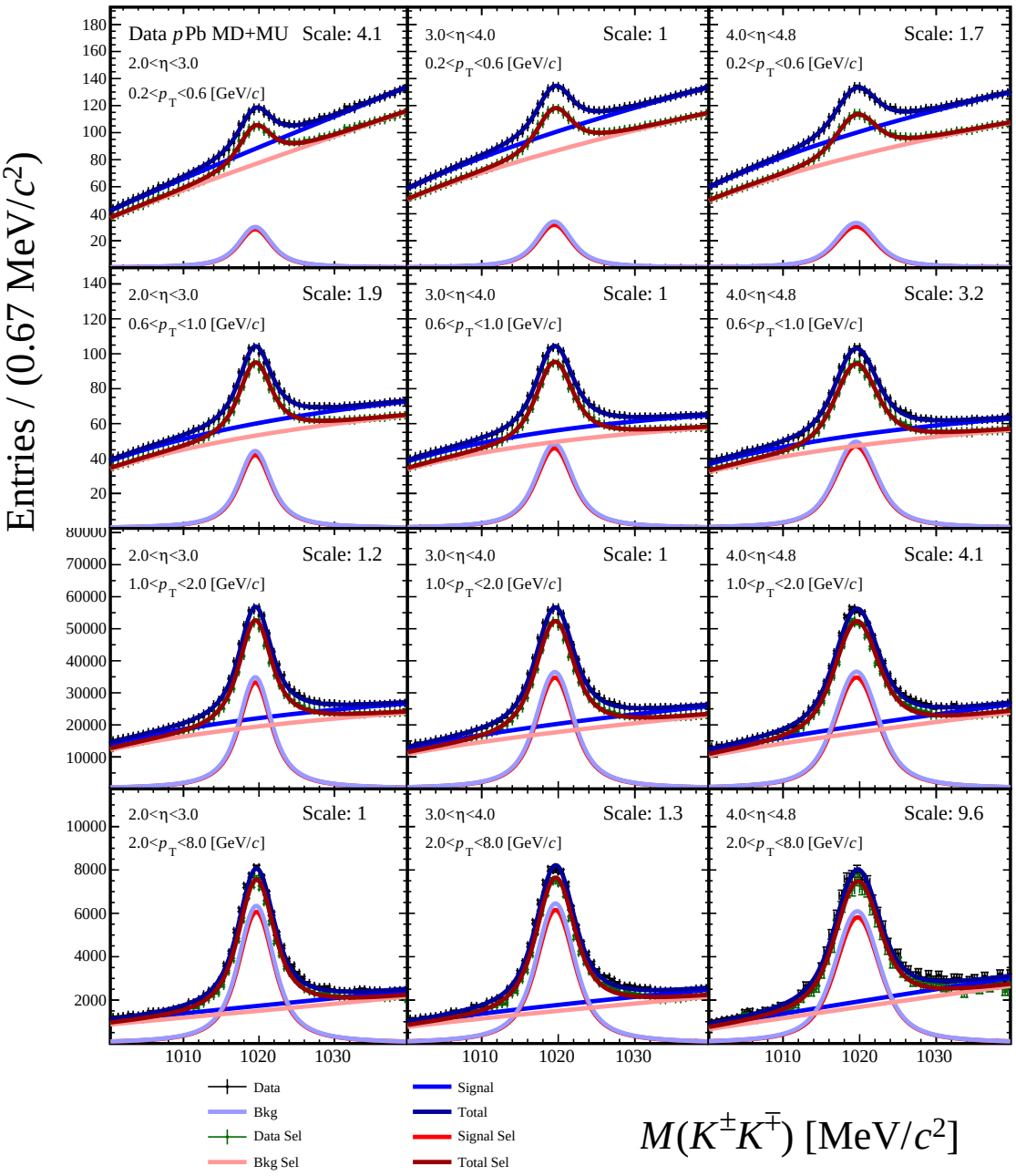


Figure B.12: Signal extraction of $\phi \rightarrow K^{\pm}K^{\mp}$ for data sample pPb collisions for the different kinematic bins (η, p_{T}). The (η, p_{T}) values are indicated in the top left side of each plot. Some plots are re-scaled to make them appear in the centre of the plot, the scale factor is indicated in the top right side.

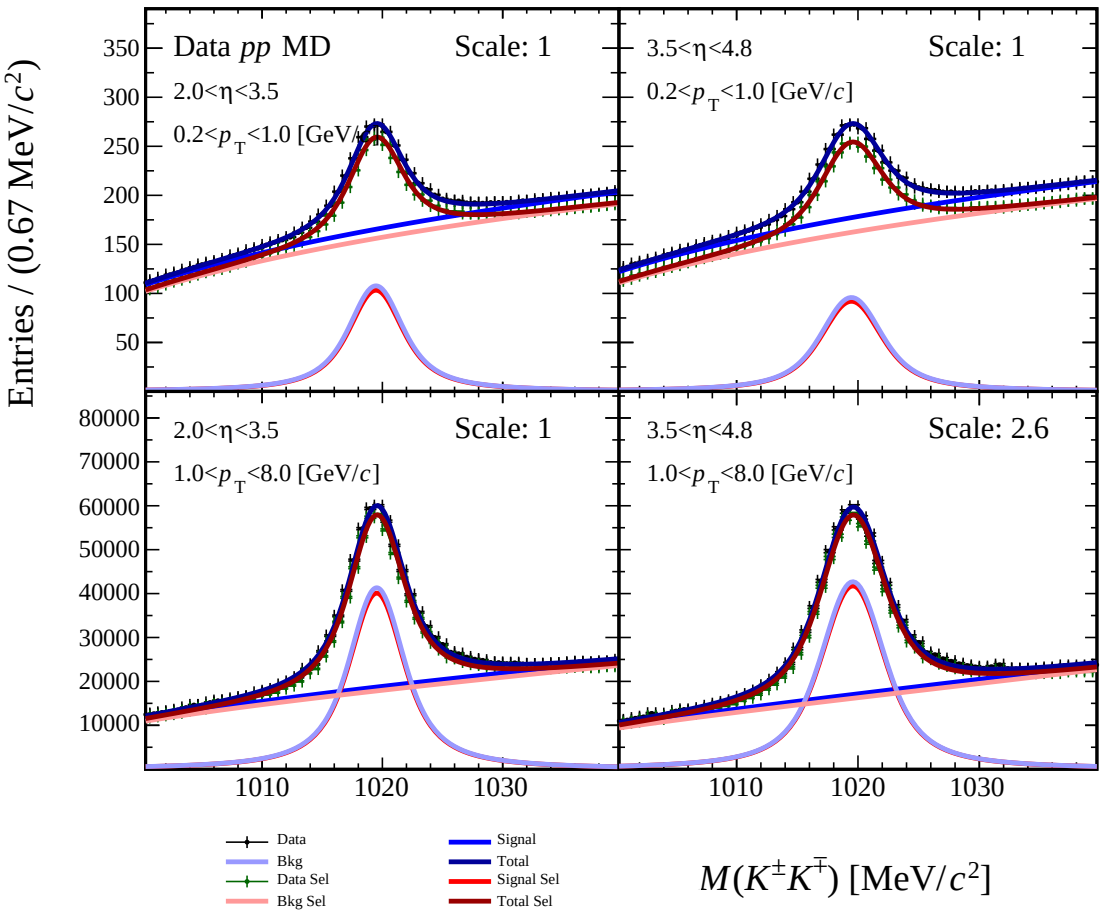


Figure B.13: Signal extraction of $\phi \to K^{\pm}K^{\mp}$ for data sample of pp collisions for the different kinematic bins (η, p_{T}). The (η, p_{T}) values are indicated in the top left side of each plot. Some plots are re-scaled to make them appear in the centre of the plot, the scale factor is indicated in the top right side.

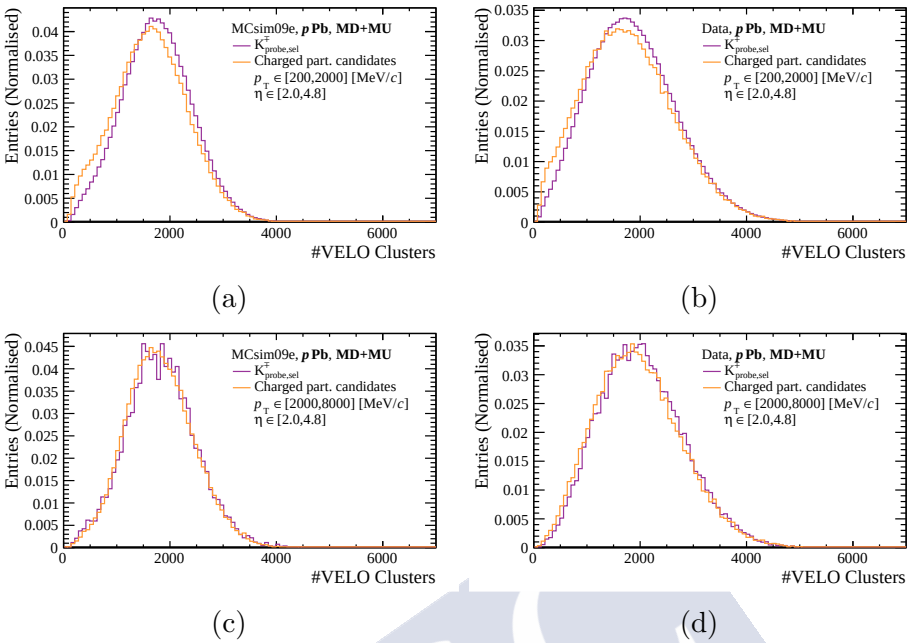


Figure B.14: Comparison of $N_{\text{VELO}}^{\text{clusters}}$ for charged particle candidates and $K^{\pm}$ probes in the pPb configuration for simulation (left) and data (right) for $200 < p_{\text{T}} < 2000\,\text{MeV}/c$ (top) and $2000 < p_{\text{T}} < 8000\,\text{MeV}/c$ (bottom).

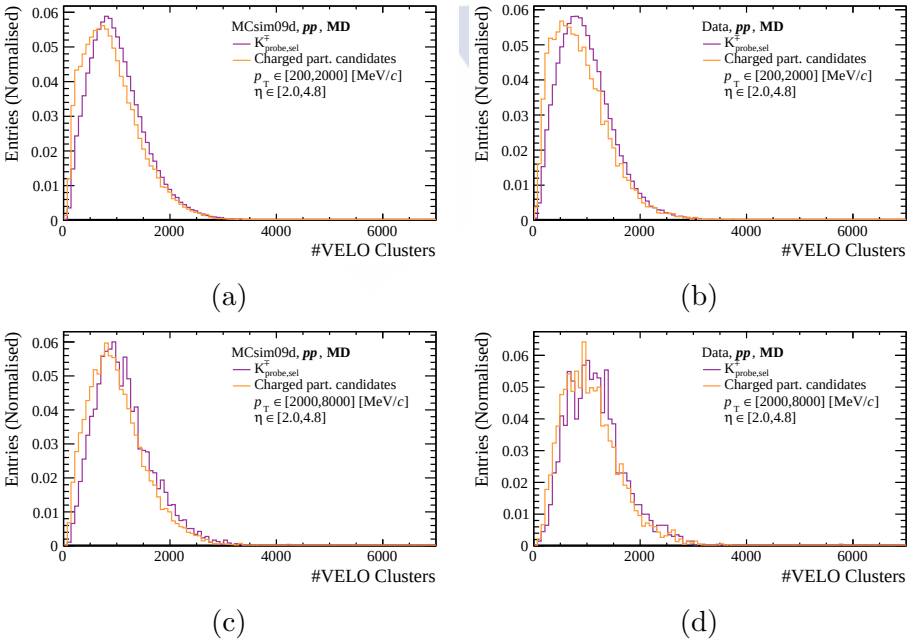


Figure B.15: Comparison of $N_{\text{VELO}}^{\text{clusters}}$ for charged particle candidates and $K^{\pm}$ probes in the pp configuration for simulation (left) and data (right) for $200 < p_{\text{T}} < 2000\,\text{MeV}/c$ (top) and $2000 < p_{\text{T}} < 8000\,\text{MeV}/c$ (bottom).

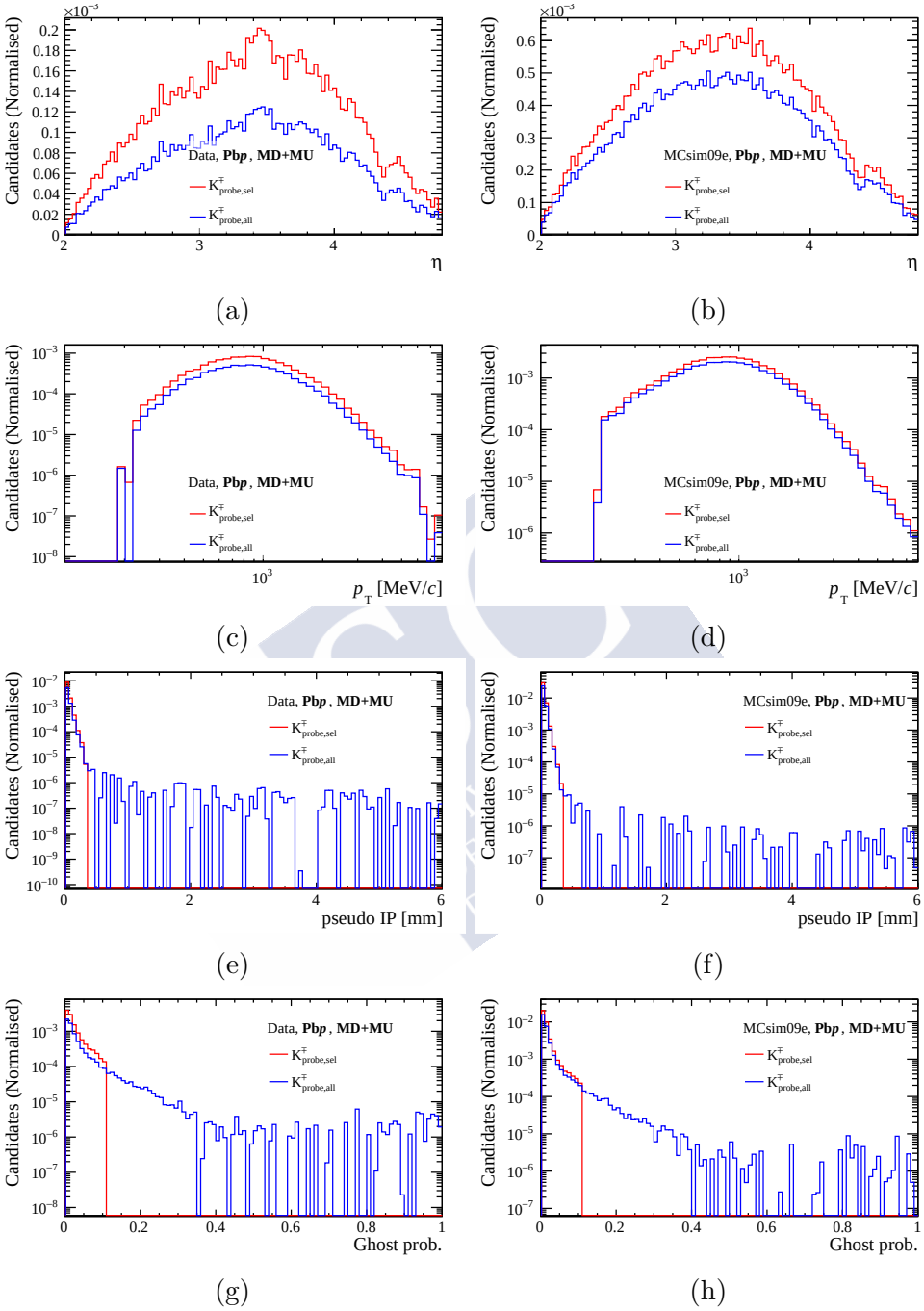


Figure B.16: Distribution of K^-_{probe} with respect to η, p_{T}, `pseudoIP` and `GhostP`, before (blue line) and after (red line) applying the selection requirements for the Pbp dataset. The background contribution, visible in Fig. 6.13, has been extracted using the sPlot technique [172].

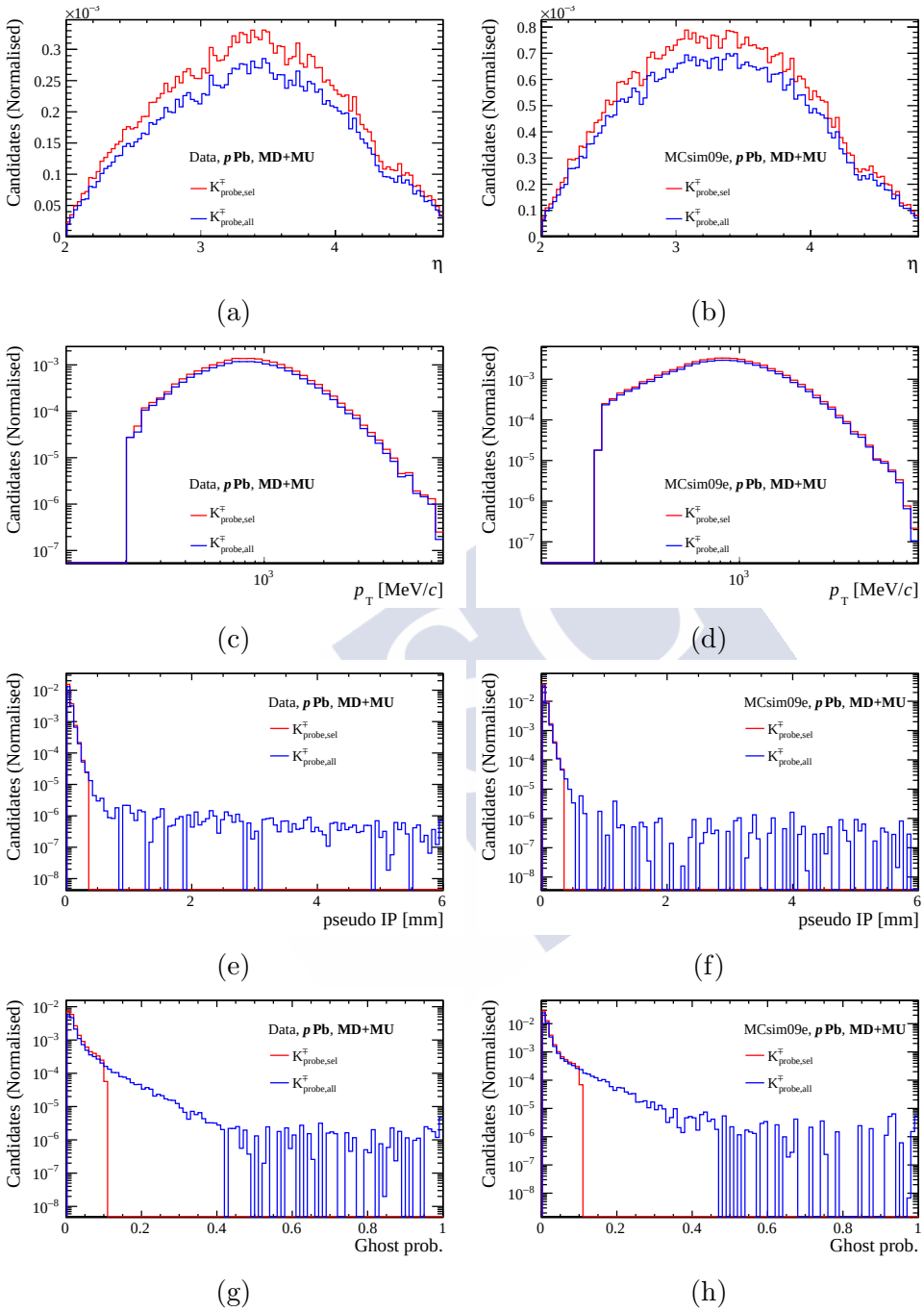


Figure B.17: Distribution of K^-_{probe} with respect to η, p_{T}, `pseudoIP` and `GhostP`, before (blue line) and after (red line) applying the selection requirements for the pPb dataset. The background contribution, visible in Fig. 6.13, has been extracted using the sPlot technique [172].

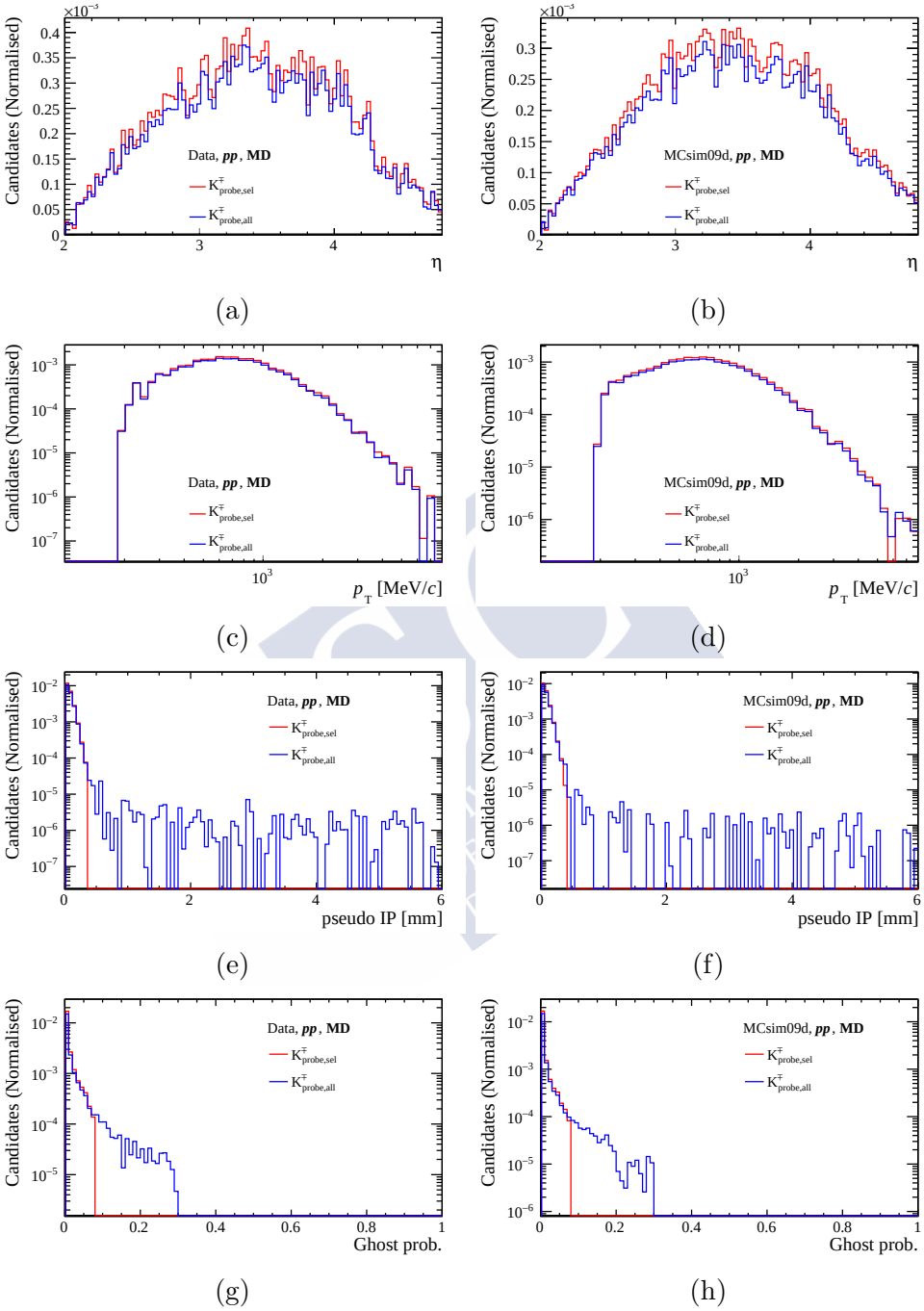


Figure B.18: Distribution of K^-_{probe} with respect to η, p_{T}, `pseudoIP` and `GhostP`, before (blue line) and after (red line) applying the selection requirements for the pp dataset. The background contribution, visible in Fig. 6.13, has been extracted using the sPlot technique [172].

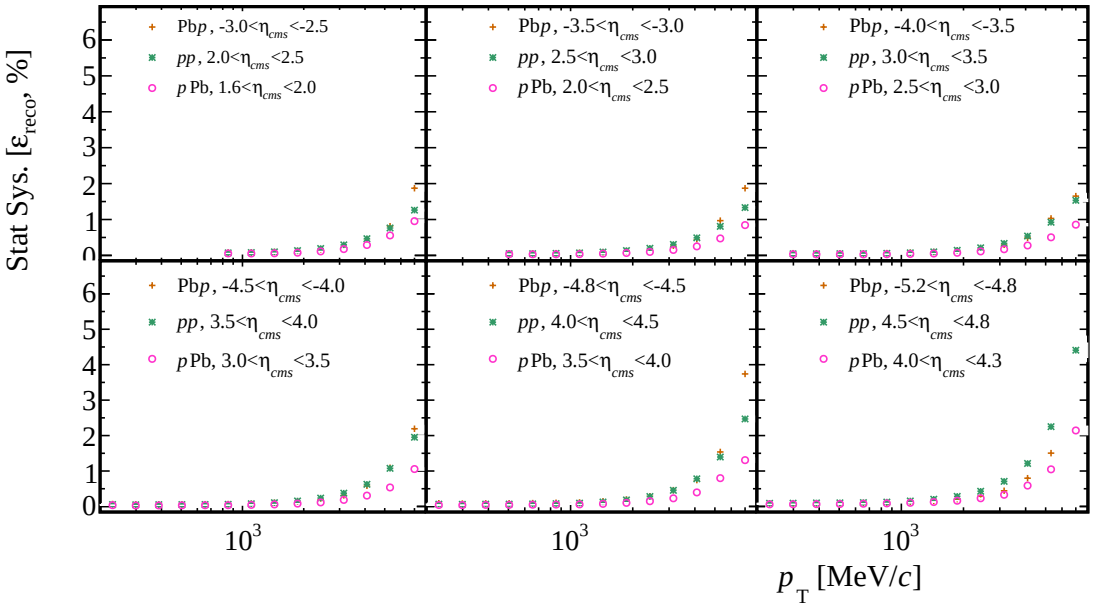


Figure B.19: Relative uncertainty in $\varepsilon_{\text{reco}}$ for Pbp, pPb and pp data due to the limited size of the simulation sample.

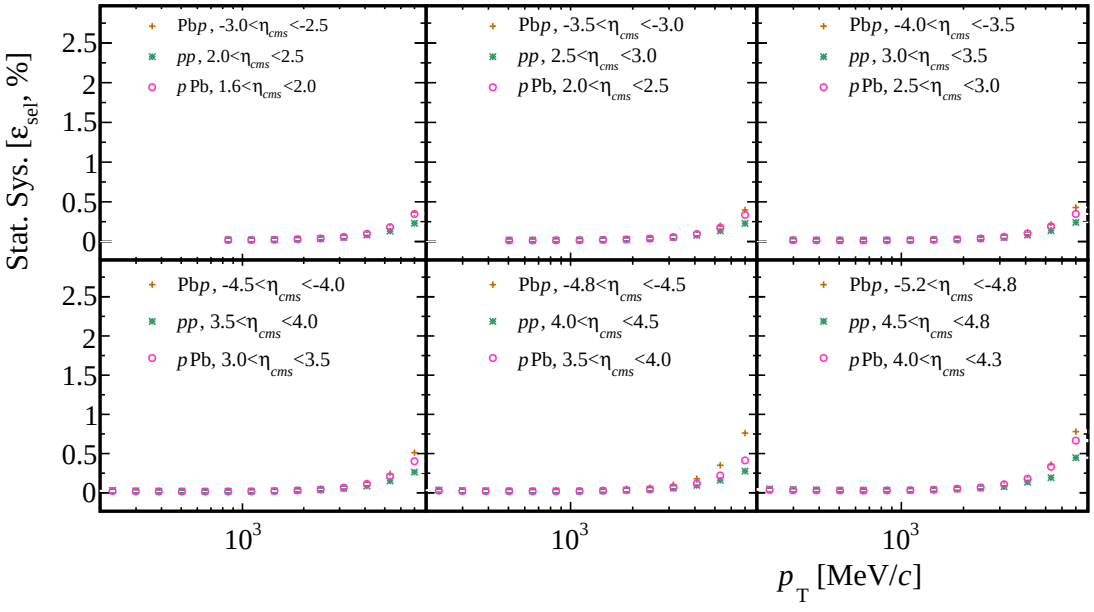


Figure B.20: Relative uncertainty in ε_{sel} for Pbp, pPb and pp data due to the limited size of the simulation sample.

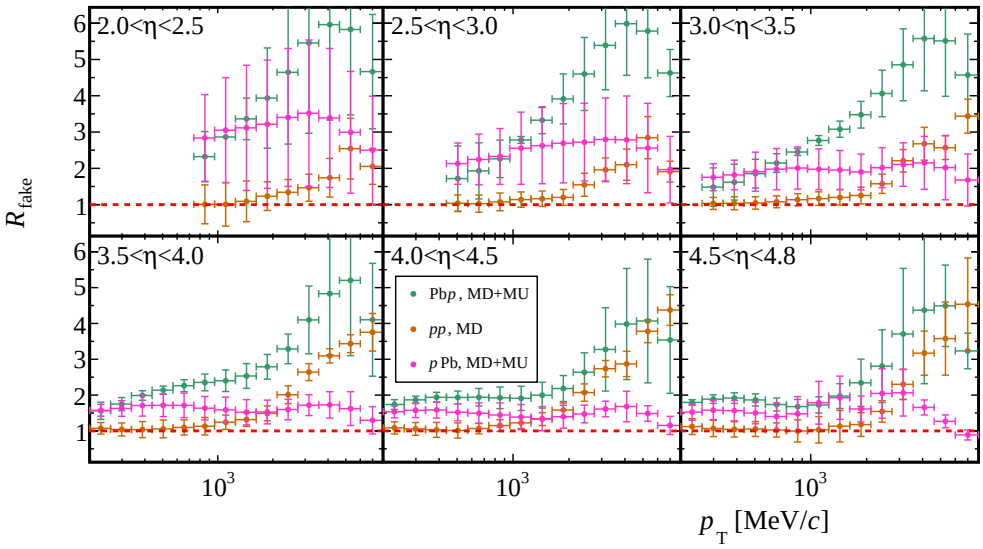


Figure B.21: Discrepancy data over simulation R_{fake} for fake tracks in Pbp, pPb and pp data sets. **Simulation is not weighted for the detector occupancy**. The values have been obtained using Eq. 6.17 and with the fake track proxy requirements of Tab. 6.4. Error bars include statistical and systematic uncertainty added quadratically.

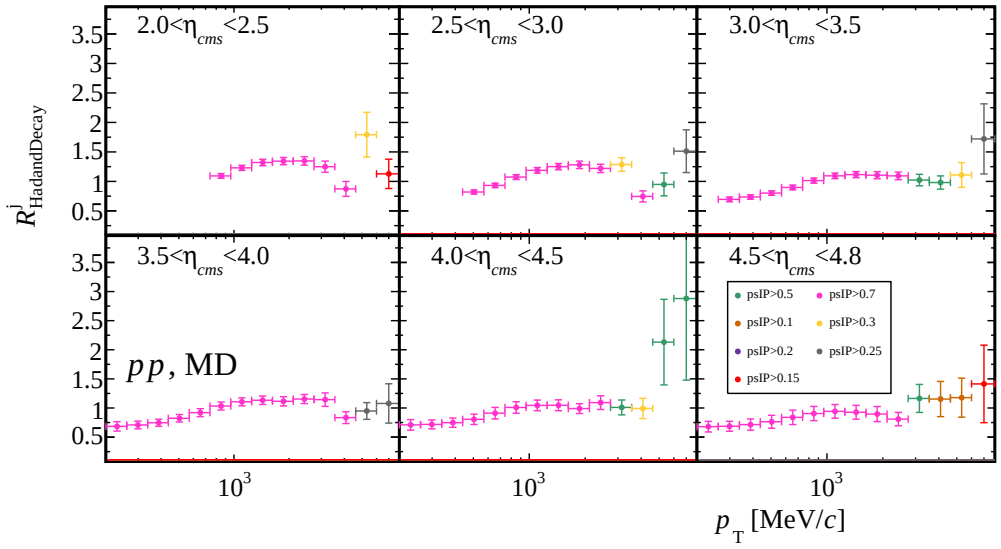


Figure B.22: $R_{\text{had\&decays}}$ for pp computed with Eq. 6.17. The chosen requirement in `pseudoIP` is showed in the legend.

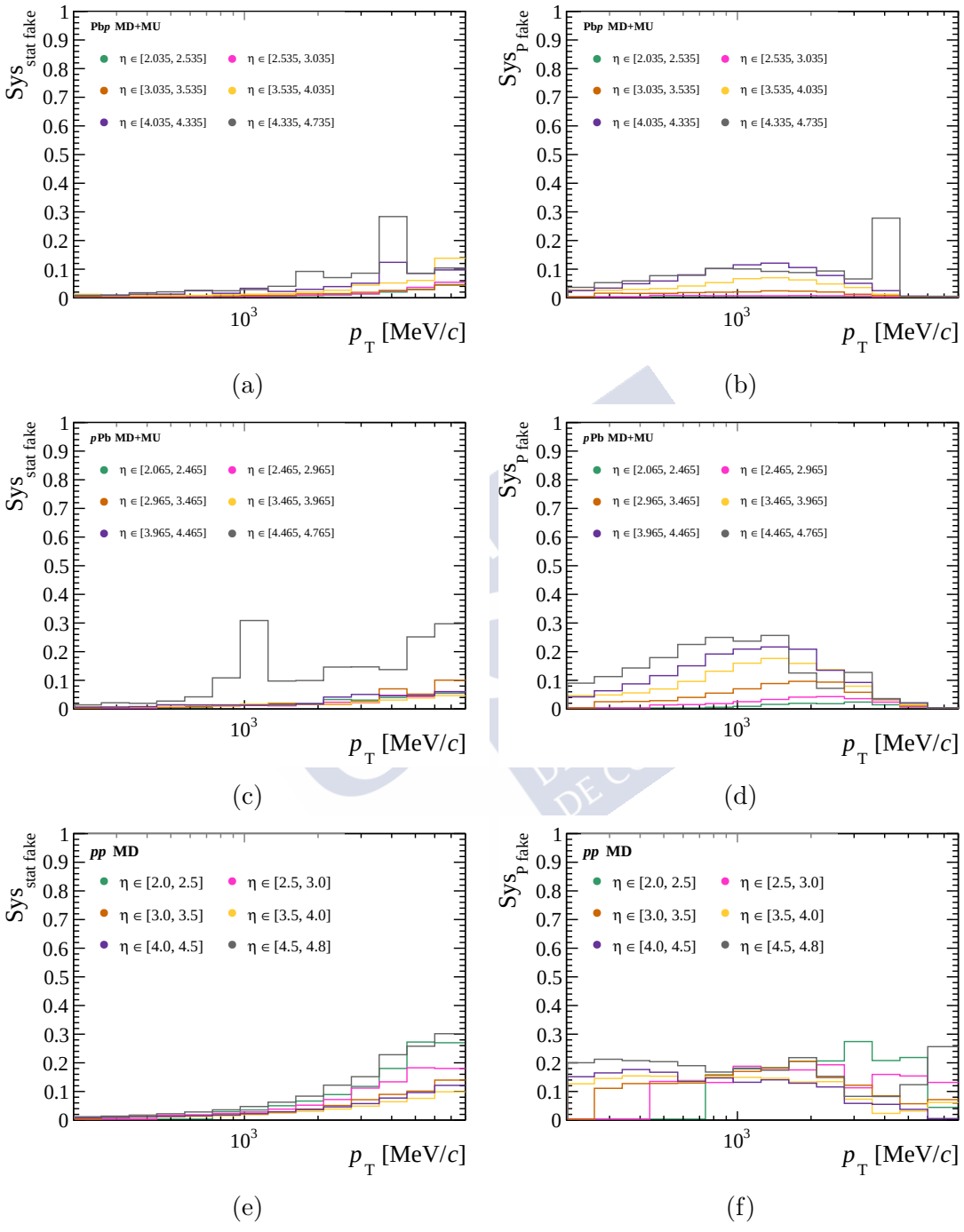


Figure B.23: Relative uncertainty on R_{fake} from sample size (left) and from fake track purity (right) for Pbp, pPb and pp samples.

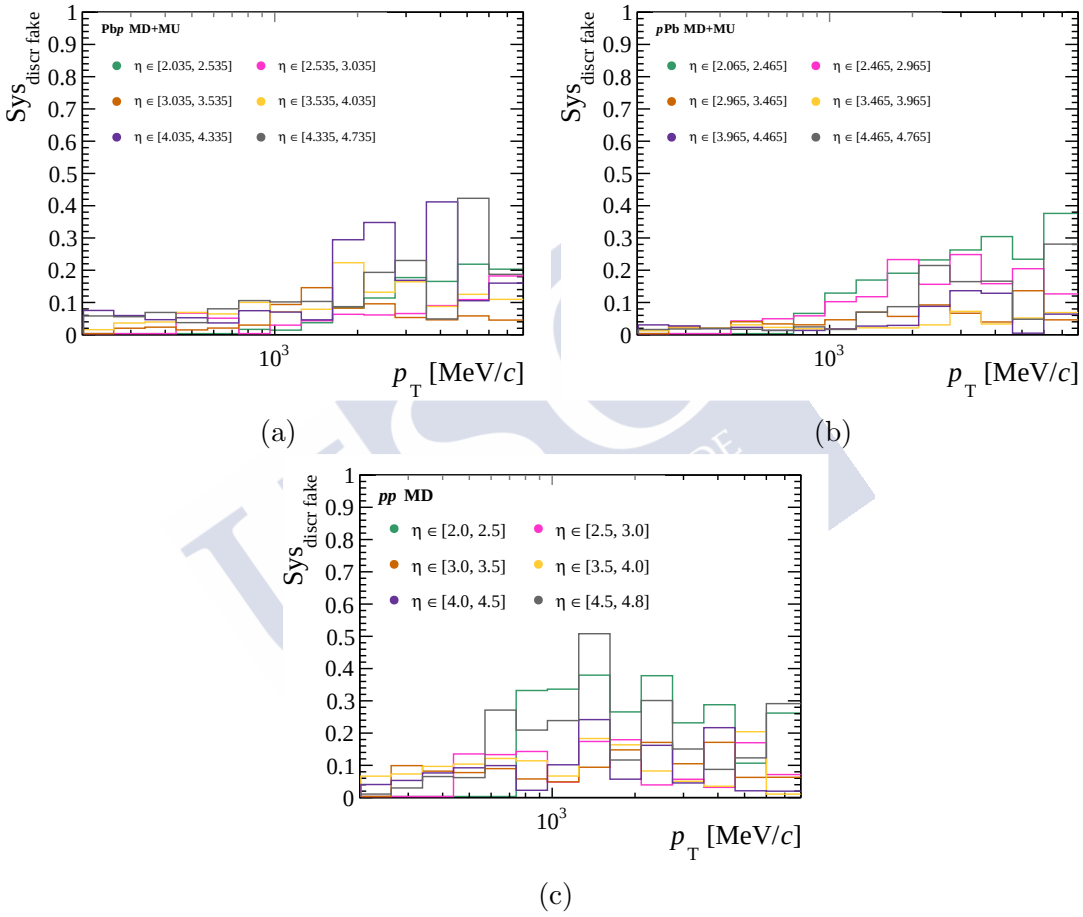
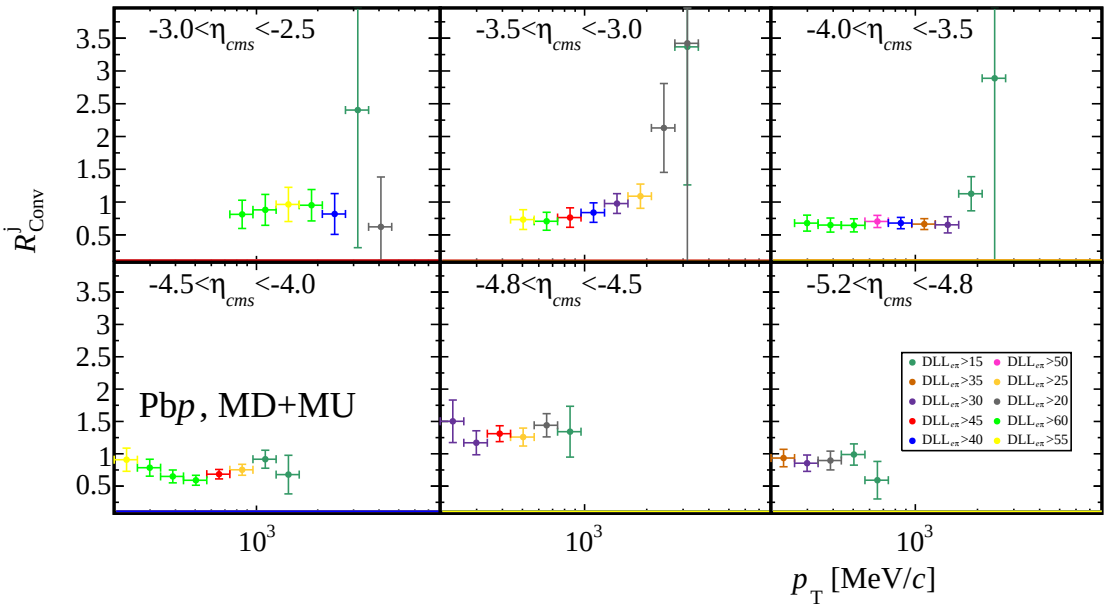
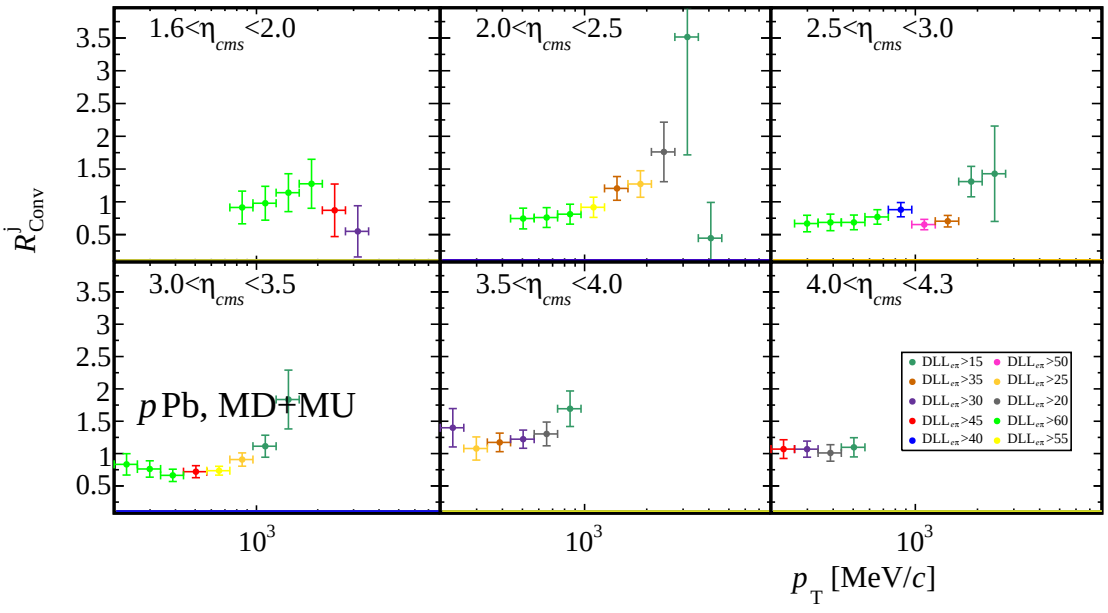


Figure B.24: Relative uncertainty on R_{fake} from discrepancy with alternative proxy (right) for Pbp, pPb and pp samples.



(a)



(b)

Figure B.25: R_γ for Pbp and pPb computed with Eq. 6.17. The chosen requirement in $\mathrm{DLL}_{e\pi}$ is showed in the legend.

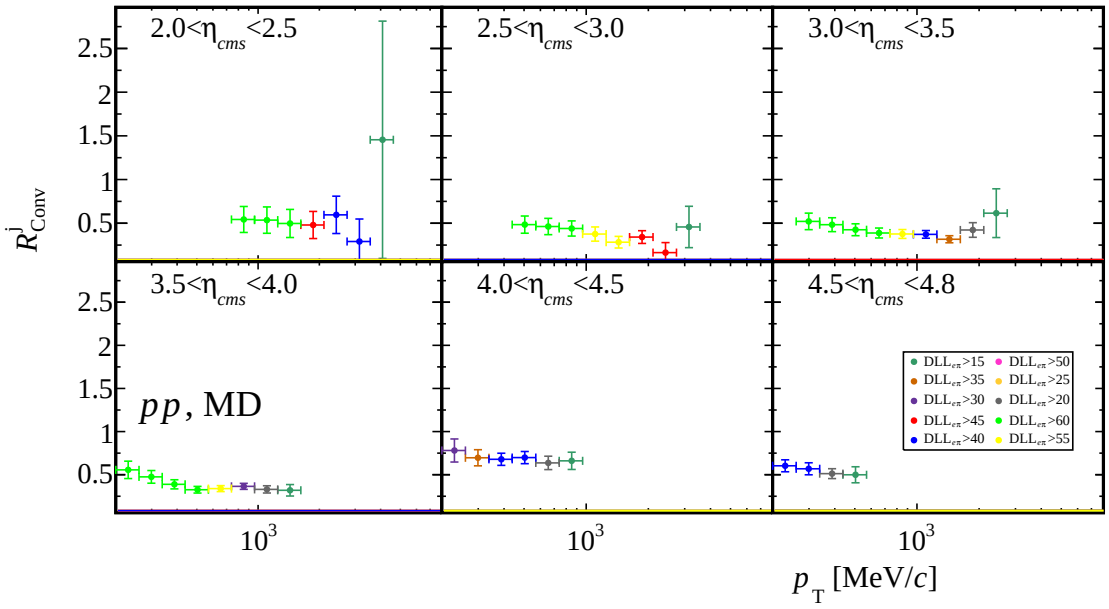


Figure B.26: R_γ for pp computed with Eq. 6.17. The chosen requirement in $\mathrm{DLL}_{e\pi}$ is showed in the legend.

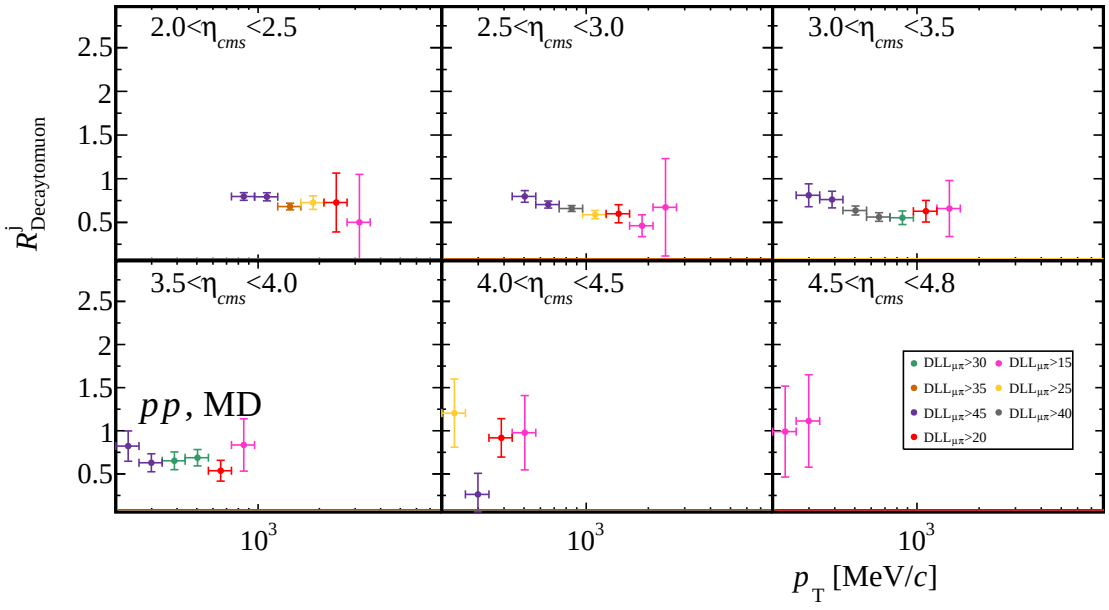
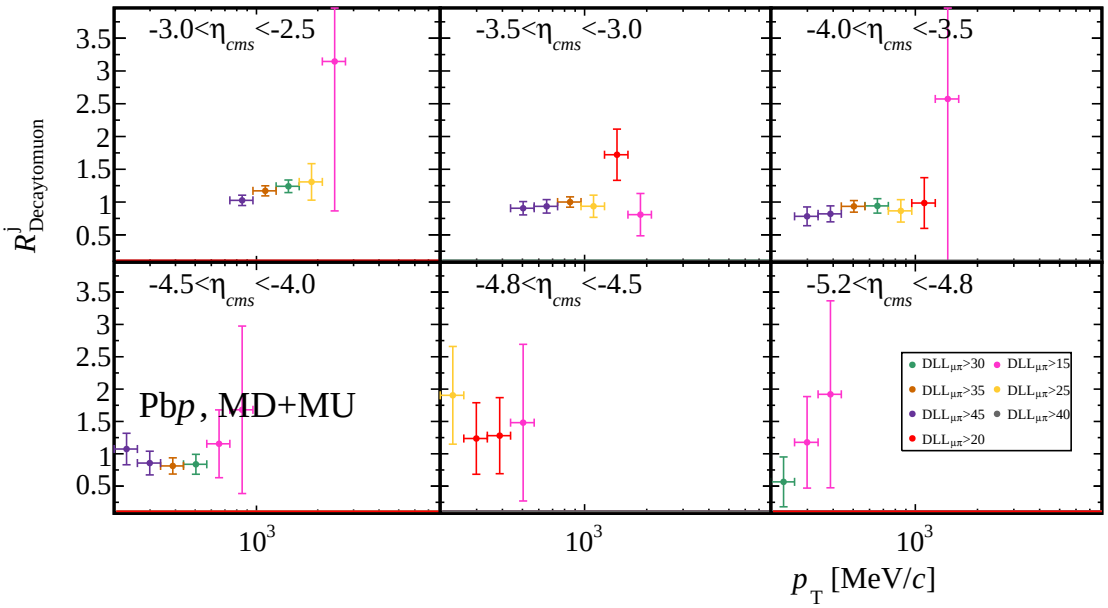
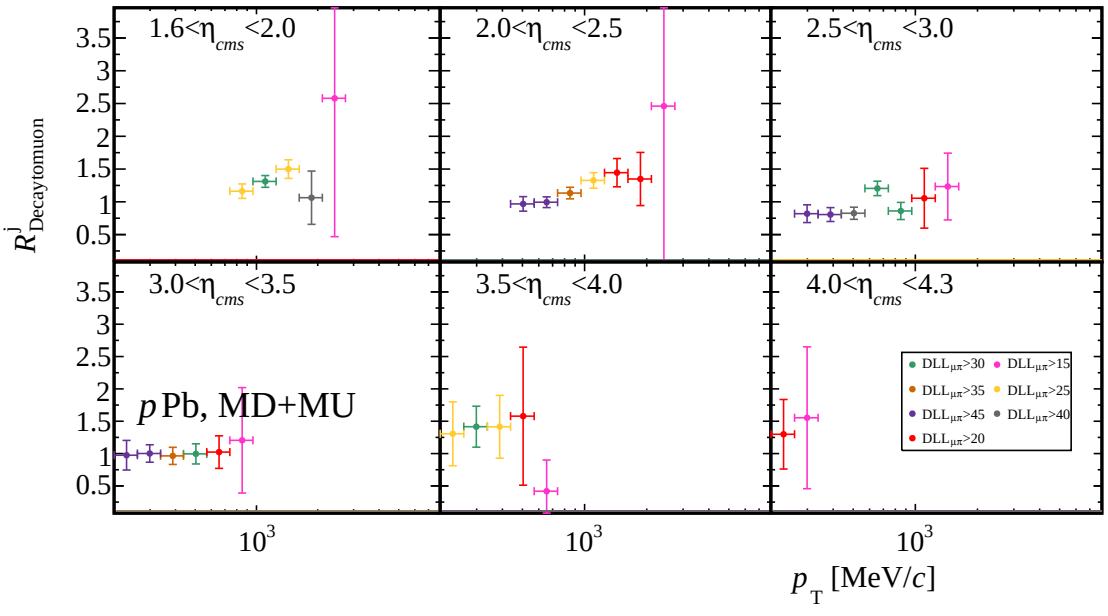


Figure B.27: $R_{X\to\mu}$ for pp computed with Eq. 6.17. The chosen requirement in $\mathrm{DLL}_{\mu\pi}$ is showed in the legend.

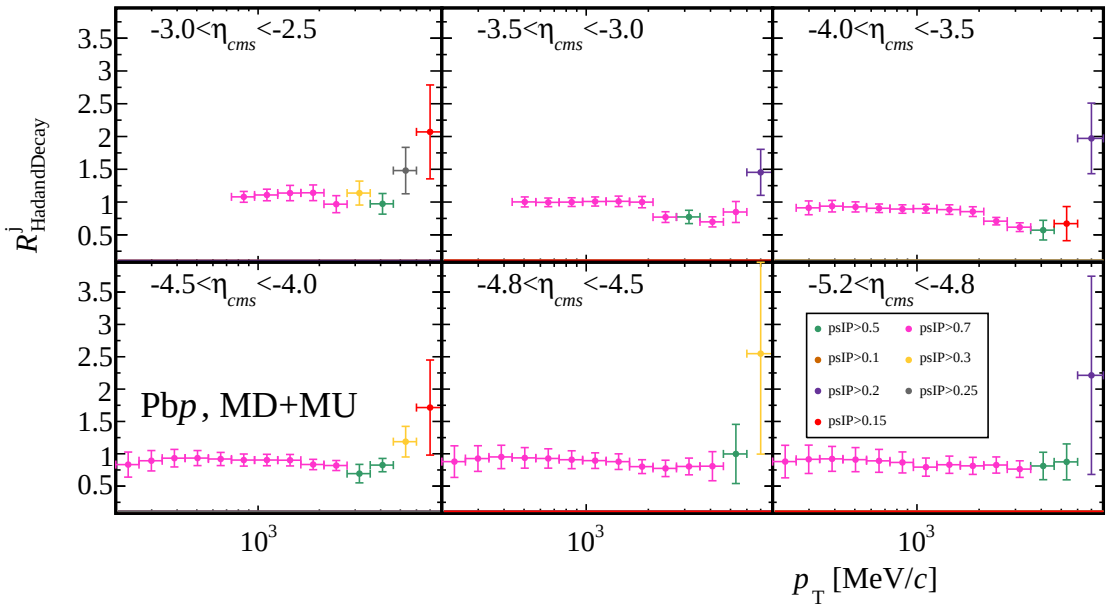


(a)

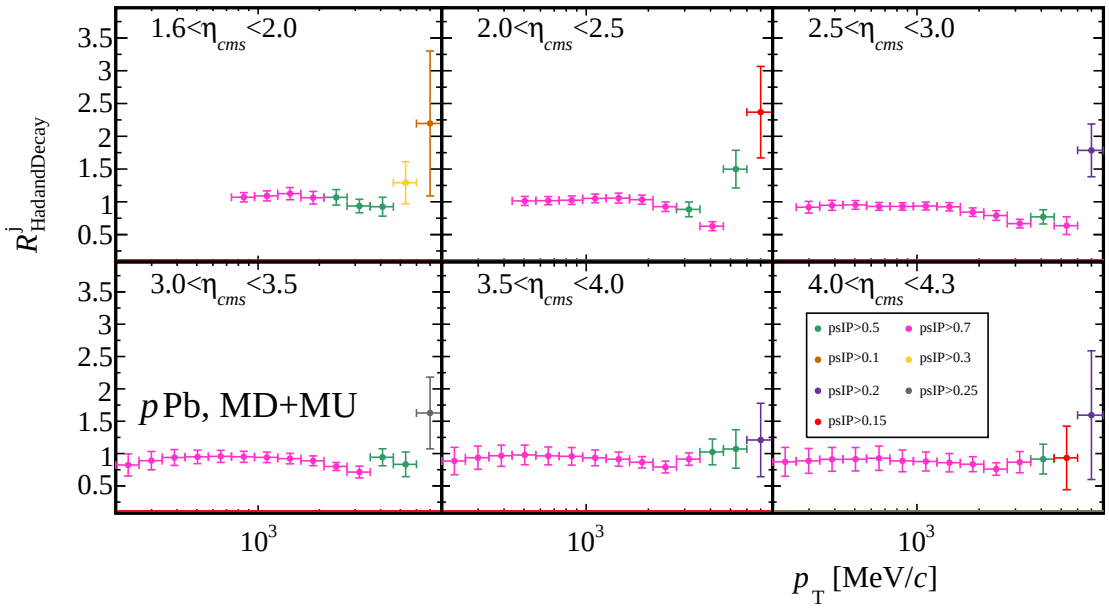


(b)

Figure B.28: $R_{X\to\mu}$ for Pbp and pPb computed with Eq. 6.17. The chosen requirement in $\mathrm{DLL}_{\mu\pi}$ is showed in the legend.



(a)



(b)

Figure B.29: $R_{Had\&Decay}$ for Pbp and pPb computed with Eq. 6.17. The chosen requirement in `pseudoIP` is showed in the legend.

List of Figures

List of Tables

Asunto: Re: Reuse Figure from PDG2020
De: PDG PDG <pdg@lbl.gov>
Fecha: 29/6/21 18:49
Para: Óscar Boente <oscar.boente@usc.es>

Dear Óscar,

Yes, you have the permission of the authors to reproduce
our figures, as long as you give credit to:
P.A. Zyla et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020) and 2021 update.

Thank you and good luck in your career!

Best regards,
Michael Barnett

On Tue, Jun 29, 2021 at 1:59 AM Óscar Boente <oscar.boente@usc.es> wrote:

> Dear Sir or Madam,
>
> My name is Óscar Boente, I am an experimental particle physics PhD student in the process of finishing my PhD thesis at the University of Santiago de Compostela (Spain). I would like to reuse one of the figures published in the PDG2020 review in the introductory chapter. Could you grant this permission?
>
> The figure to be reused is in Chapter 9 (Quantum Chromodynamics), figure 9.3 (summary of alpha_s measurements). The title of my thesis is "Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment". Thanks in advance for your help.
>
> Kind Regards,
>
> Óscar Boente
>
> --
> Óscar Boente García PhD student
> Instituto Galego de Física de Altas Enerxías - Universidade de Santiago de Compostela
>
> Rúa de Xoaquín Díaz de Rábago s/n, Campus Vida - Universidade de Santiago de Compostela, Santiago de Compostela, 15782, Spain
>
> IGFAE: oscar.boente@usc.es Tel: +34 8818 13983
> CERN: oscar.boente@cern.ch Tel: +41 22 76 76585

CCC Marketplace™

This is a License Agreement between Óscar Boente García ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the CCC Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the CCC Terms and Conditions below.

Order Date	29-Jun-2021	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1129421-1		
ISSN	1545-4134	**Publisher**	ANNUAL REVIEWS
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Annual review of nuclear and particle science	**Rightsholder**	Annual Reviews, Inc.
		Publication Type	e-Journal
Date	01/01/1978	**URL**	http://arjournals.annualreviews.org/loi/nucl
Language	English		
Country	United States of America		

REQUEST DETAILS

Portion Type	Image/photo/illustration	**Distribution**	Worldwide
Number of images / photos / illustrations	1	**Translation**	Original language of publication
Format (select all that apply)	Print, Electronic	**Copies for the disabled?**	No
		Minor editing privileges?	No
Who will republish the content?	Academic institution	**Incidental promotional use?**	No
Duration of Use	Life of current edition	**Currency**	USD
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment	**Institution name**	Universidade de Santiago de Compostela
		Expected presentation date	2021-09-10
Instructor name	Óscar Boente García		

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	Óscar Boente García

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 1	**Title of the article/chapter the portion is from**	1. INTRODUCTION
Editor of portion(s)	N/A	**Author of portion(s)**	Wit Busza, Krishna Rajagopal and Wilke van der Schee
Volume of serial or monograph	68		
Page or page range of portion	339-376	**Issue, if republishing an article from a serial**	N/A
		Publication date of portion	2018-10-19

CCC Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

    3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on

paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated

by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

CCC Marketplace™

This is a License Agreement between Óscar Boente García ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the CCC Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the CCC Terms and Conditions below.

Order Date	29-Jun-2021	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1129423-1		
ISSN	1545-4134	**Publisher**	ANNUAL REVIEWS
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Annual review of nuclear and particle science	**Rightsholder**	Annual Reviews, Inc.
		Publication Type	e-Journal
Date	01/01/1978	**URL**	http://arjournals.annualreviews.org/loi/nucl
Language	English		
Country	United States of America		

REQUEST DETAILS

Portion Type	Image/photo/illustration	**Distribution**	Worldwide
Number of images / photos / illustrations	1	**Translation**	Original language of publication
Format (select all that apply)	Print, Electronic	**Copies for the disabled?**	No
		Minor editing privileges?	No
Who will republish the content?	Academic institution	**Incidental promotional use?**	No
Duration of Use	Life of current edition	**Currency**	USD
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment	**Institution name**	Universidade de Santiago de Compostela
		Expected presentation date	2021-09-10
Instructor name	Abraham Gallas Torreira		

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	Óscar Boente García

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 2	**Title of the article/chapter the portion is from**	The Color Glass Condensate, chapter 1
Editor of portion(s)	N/A	**Author of portion(s)**	Francois Gelis, Edmond Iancu, Jamal Jalilian-Marian, and Raju Venugopalan
Volume of serial or monograph	60		
Page or page range of portion	463-489	**Issue, if republishing an article from a serial**	N/A
		Publication date of portion	2010-11-01

CCC Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

    3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by

removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any

unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

Asunto: Fwd: Fw: Request Figure Reuse JHEP02(2009)007
De: JHEP Editorial Office <jhep-eo@jhep.sissa.it>
Fecha: 30/6/21 11:28
Para: oscar.boente@usc.es
CC: jhep-eo@jhep.sissa.it

Dear Dr. Boente,

IoP have forwarded your message to us at JHEP Editorial Office as we deal with permissions.

We hereby grant you permission to reproduce the figure mentioned in your message, provided the original source of publication is properly credited, also by means of the DOI.

--
Best regards,

Maria Teresa Leo
JHEP Editorial Office

From: Óscar Boente <oscar.boente@usc.es>
Sent: 29 June 2021 10:40
To: Permissions <permissions@ioppublishing.org>
Subject: Request Figure Reuse JHEP02(2009)007

Dear Sir or Madam,

My name is Óscar Boente, I am an experimental particle physics PhD student in the process of finishing my PhD thesis at the University of Santiago de Compostela (Spain). I would like to reuse one of the figures published in the article: T. Gleisberg et al JHEP02(2009)007. Could you grant this permission?

The figure to be reused is Figure 1 (Pictorial representation of a ttbarh event as produced by an event generator). The title of my thesis is "Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment". Thanks in advance for your help.

Kind Regards,

Óscar Boente

--
Óscar Boente García PhD student
Instituto Galego de Física de Altas Enerxías - Universidade de Santiago de Compostela

Rúa de Xoaquín Díaz de Rábago s/n, Campus Vida - Universidade de Santiago de Compostela, Santiago de Compostela, 15782, Spain

IGFAE: oscar.boente@usc.es Tel: +34 8818 13983
CERN: oscar.boente@cern.ch Tel: +41 22 76 76585

Asunto: Fwd: Fw: Request Figure Reuse JHEP02(2009)007
De: JHEP Editorial Office <jhep-eo@jhep.sissa.it>
Fecha: 30/6/21 11:28
Para: oscar.boente@usc.es
CC: jhep-eo@jhep.sissa.it

Dear Dr. Boente,

IoP have forwarded your message to us at JHEP Editorial Office as we deal with permissions.

We hereby grant you permission to reproduce the figure mentioned in your message, provided the original source of publication is properly credited, also by means of the DOI.

--
Best regards,

Maria Teresa Leo
JHEP Editorial Office

From: Óscar Boente <oscar.boente@usc.es>
Sent: 29 June 2021 10:40
To: Permissions <permissions@ioppublishing.org>
Subject: Request Figure Reuse JHEP02(2009)007

Dear Sir or Madam,

My name is Óscar Boente, I am an experimental particle physics PhD student in the process of finishing my PhD thesis at the University of Santiago de Compostela (Spain). I would like to reuse one of the figures published in the article: T. Gleisberg et al JHEP02(2009)007. Could you grant this permission?

The figure to be reused is Figure 1 (Pictorial representation of a ttbarh event as produced by an event generator). The title of my thesis is "Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment". Thanks in advance for your help.

Kind Regards,

Óscar Boente

--
Óscar Boente García PhD student
Instituto Galego de Física de Altas Enerxías - Universidade de Santiago de Compostela

Rúa de Xoaquín Díaz de Rábago s/n, Campus Vida - Universidade de Santiago de Compostela, Santiago de Compostela, 15782, Spain

IGFAE: oscar.boente@usc.es Tel: +34 8818 13983
CERN: oscar.boente@cern.ch Tel: +41 22 76 76585

Asunto: Re: Reuse figure from: The LHCb Collaboration et al 2008 JINST 3 S08005
De: Permissions <permissions@ioppublishing.org>
Fecha: 30/6/21 10:13
Para: Óscar Boente <oscar.boente@usc.es>
CC: jinst-eo <jinst-eo@jinst.sissa.it>

Dear Óscar Boente,

Thank you for your request to reproduce material published by IOP Publishing *in your thesis, **"Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment"***

Regarding:

- ***Figures 2.1, 5.35, 6.1, 10.1, & 10.2** from **"The LHCb Detector at the LHC"***

We are happy to grant permission for the use you request on the terms set out below.

License to publish material published by IOP Publishing

Please provide the below to your new publisher as proof of permission.

Conditions

Non-exclusive, non-transferrable, revocable, worldwide, permission to use the material in print and electronic form will be granted **subject to the following conditions:**

- Permission will be cancelled without notice if you fail to fulfil any of the conditions of this letter.
- You will make reasonable efforts to contact the author(s) to seek consent for your intended use. Contacting one author acting expressly as authorised agent for their co-authors is acceptable.
- You will reproduce the following prominently alongside the material:
  - the source of the material, including author, article title, title of journal, volume number, issue number (if relevant), page range (or first page if this is the only information available) and date of first publication. This information can be contained in a footnote or reference note; or
  - a link back to the article (via DOI); and
  - ***if practical and IN ALL CASES for works published under any of the Creative Commons licences the words "© IOP Publishing Ltd and SISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved"***
- The material will not, without the express permission of the author(s), be used in any way which, in the opinion of IOP Publishing, could distort or alter the author(s)' original intention(s) and meaning, be prejudicial to the honour or reputation of the author(s) and/or imply endorsement by the author(s) and/or IOP Publishing and/or SISSA Medialab Srl.
- Payment of £0 is received in full by IOP Publishing prior to use.

This permission does not apply to any material/figure which is credited to another source in our publication or has been obtained from a third party. Express permission for such materials/figures must be obtained from the copyright owner.

Kind regards,

Sophie
Copyright & Permissions Team
Sophie Brittain - Rights & Permissions Assistant
Cameron Wood - Legal & Rights Adviser
Contact Details
E-mail: permissions@ioppublishing.org

For further information about copyright and how to request permission:
https://publishingsupport.iopscience.iop.org/copyright-journals/
See also: https://publishingsupport.iopscience.iop.org/
Please see our Author Rights Policy https://publishingsupport.iopscience.iop.org/author-rights-policies/
Please note: We do not provide signed permission forms as a separate attachment. Please print this email and provide it to your publisher as proof of permission. **Please note**: Any statements made by IOP Publishing to the effect that authors do not need to get permission to use any content where IOP Publishing is not the publisher is not intended to constitute any sort of legal advice. Authors must make their own decisions as to the suitability of the content they are using and whether they require permission for it to be published within their article.

From: Óscar Boente <oscar.boente@usc.es>
Sent: 29 June 2021 11:27
To: Permissions <permissions@ioppublishing.org>
Subject: Reuse figure from: The LHCb Collaboration et al 2008 JINST 3 S08005

Dear Sir or Madam,
My name is Óscar Boente, I am an experimental particle physics PhD student in the process of finishing my PhD thesis at the University of Santiago de Compostela (Spain). I would like to reuse some figures in my PhD thesis from the following publication:
Journal of Instrumentation, Volume 3, August 2008: The LHCb Detector at the LHC
The LHCb Collaboration et al, 2008 JINST 3 S08005
The figures to be reused are:
- Figure 2.1
- Figure 5.35
- Figure 6.1
- Figure 10.1
- Figure 10.2

Could you grant permission for being reused? The title of the thesis is "Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment". Thanks in advance for your help.
Kind Regards,
Óscar Boente
--
Óscar Boente García PhD student
Instituto Galego de Física de Altas Enerxías - Universidade de Santiago de Compostela

Rúa de Xoaquín Díaz de Rábago s/n, Campus Vida - Universidade
de Santiago de Compostela, Santiago de Compostela, 15782, Spain

IGFAE: oscar.boente@usc.es Tel: +34 8818 13983
CERN: oscar.boente@cern.ch Tel: +41 22 76 76585

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jun 29, 2021

This Agreement between Oscar Boente Garcia ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5098121148949
License date	Jun 29, 2021
Licensed Content Publisher	Elsevier
Licensed Content Publication	Physics Letters B
Licensed Content Title	Exponential hadronic spectrum and quark liberation
Licensed Content Author	N. Cabibbo,G. Parisi
Licensed Content Date	Oct 13, 1975
Licensed Content Volume	59
Licensed Content Issue	1
Licensed Content Pages	3
Start Page	67

End Page	69
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Analysis of charged particle production in proton-nucleus and proton-proton collisions at the LHCb experiment
Institution name	Universidade de Santiago de Compostela
Expected presentation date	Sep 2021
Portions	Fig.1 (Schematic phase diagram of hadronic matter) from Phys.Lett.B 59 (1975), 67-69, DOI: 10.1016/0370-2693(75)90158-6,
Requestor Location	Oscar Boente Garcia Camino cachadas adega 12 Ponteareas, Pontevedra 36860 Spain Attn: Oscar Boente Garcia
Publisher Tax ID	GB 494 6272 12

Total 0.00 EUR

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not

allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the

work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.10

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

REFERENCES

[1] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. **B47** (1973) 365. (Cited on pages 1 and 4.)

[2] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, arXiv:hep-ph/0409313. (Cited on pages 1 and 6.)

[3] CTEQ collaboration, R. Brock *et al.*, *Handbook of perturbative QCD; Version 1.1: September 1994*, . (Cited on pages 1 and 6.)

[4] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005. (Cited on pages 1, 29, 32, 37, 38, 40, 209, and 244.)

[5] HiRes, MIA collaborations, T. Abu-Zayyad *et al.*, *Evidence for Changing of Cosmic Ray Composition between 10**17-eV and 10**18-eV from Multicomponent Measurements*, Phys. Rev. Lett. **84** (2000) 4276, arXiv:astro-ph/9911144. (Cited on pages 2, 16, and 204.)

[6] KASCADE collaboration, T. Antoni *et al.*, *A Nonparametric approach to infer the energy spectrum and the mass composition of cosmic rays*, Astropart. Phys. **16** (2002) 245, arXiv:astro-ph/0102443. (Cited on pages 2, 16, and 204.)

[7] J. Albrecht *et al.*, *The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider*, arXiv:2105.06148. (Cited on pages 2, 16, 204, and 207.)

[8] A. Buckley *et al.*, *General-purpose event generators for LHC physics*, Phys. Rept. **504** (2011) 145, arXiv:1101.2599. (Cited on pages 2, 15, and 207.)

[9] W. Busza, K. Rajagopal, and W. van der Schee, *Heavy Ion Collisions: The Big Picture, and the Big Questions*, Ann. Rev. Nucl. Part. Sci. **68** (2018) 339, arXiv:1802.04801. (Cited on pages 2, 9, 13, 14, and 244.)

[10] PHENIX collaboration, K. Adcox *et al.*, *Suppression of hadrons with large transverse momentum in central Au+Au collisions at* $\sqrt{s_{NN}}$ *= 130-GeV*, Phys. Rev. Lett. **88** (2002) 022301, arXiv:nucl-ex/0109003. (Cited on pages 2 and 14.)

[11] Y. Mehtar-Tani, J. G. Milhano, and K. Tywoniuk, *Jet physics in heavy-ion collisions*, Int. J. Mod. Phys. **A28** (2013) 1340013, arXiv:1302.2579. (Cited on pages 2 and 14.)

[12] J. L. Albacete *et al.*, *Predictions for p+Pb Collisions at sqrt s_NN = 5 TeV*, Int. J. Mod. Phys. **E22** (2013) 1330007, arXiv:1301.3395. (Cited on pages 2, 17, 18, and 208.)

[13] O. Boente García, A. A. Gallas Torreira, C. Santamarina Ríos, and R. Vázquez Gómez, *Measurement of the nuclear modification factor and prompt charged particle production in pPb, Pbp and pp collisions at* $\sqrt{s_{NN}}$*=5 TeV*, LHCb-ANA-2020-048. (Cited on page 2.)

[14] G. Altarelli, *The Standard model of particle physics*, `arXiv:hep-ph/0510281`. (Cited on page 3.)

[15] G. Altarelli, *Collider Physics within the Standard Model: a Primer*, `arXiv:1303.2842`. (Cited on page 3.)

[16] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01. (Cited on pages 3, 4, 5, and 244.)

[17] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013. (Cited on page 3.)

[18] O. Kaczmarek, F. Karsch, F. Zantow, and P. Petreczky, *Static quark anti-quark free energy and the running coupling at finite temperature*, Phys. Rev. **D70** (2004) 074505, `arXiv:hep-lat/0406036`, [Erratum: Phys.Rev.D 72, 059903 (2005)]. (Cited on page 6.)

[19] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438. (Cited on page 7.)

[20] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298. (Cited on page 7.)

[21] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641. (Cited on page 7.)

[22] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, Phys. Rept. **742** (2018) 1, `arXiv:1709.04922`. (Cited on page 7.)

[23] T.-J. Hou *et al.*, *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, Phys. Rev. **D103** (2021) 014013, `arXiv:1912.10053`. (Cited on pages 7 and 244.)

[24] S. M. Berman, J. D. Bjorken, and J. B. Kogut, *Inclusive Processes at High Transverse Momentum*, Phys. Rev. **D4** (1971) 3388. (Cited on page 7.)

[25] A. Metz and A. Vossen, *Parton Fragmentation Functions*, Prog. Part. Nucl. Phys. **91** (2016) 136, `arXiv:1607.02521`. (Cited on page 8.)

[26] D. de Florian, R. Sassot, and M. Stratmann, *Global analysis of fragmentation functions for pions and kaons and their uncertainties*, Phys. Rev. **D75** (2007) 114010, arXiv:hep-ph/0703242. (Cited on pages 8, 196, and 216.)

[27] D. de Florian *et al.*, *Parton-to-Pion Fragmentation Reloaded*, Phys. Rev. **D91** (2015) 014035, arXiv:1410.6027. (Cited on page 8.)

[28] NNPDF collaboration, V. Bertone *et al.*, *Charged hadron fragmentation functions from collider data*, Eur. Phys. J. **C78** (2018) 651, arXiv:1807.03310. (Cited on page 8.)

[29] P. Braun-Munzinger, V. Koch, T. Schäfer, and J. Stachel, *Properties of hot and dense matter from relativistic heavy ion collisions*, Phys. Rept. **621** (2016) 76, arXiv:1510.00442. (Cited on page 8.)

[30] B. Friman *et al.*, eds., *The CBM physics book: Compressed baryonic matter in laboratory experiments*, vol. 814, Springer, 2011. (Cited on page 8.)

[31] N. Cabibbo and G. Parisi, *Exponential Hadronic Spectrum and Quark Liberation*, Phys. Lett. **B59** (1975) 67. (Cited on pages 8 and 244.)

[32] Y. Aoki *et al.*, *The Order of the quantum chromodynamics transition predicted by the standard model of particle physics*, Nature **443** (2006) 675, arXiv:hep-lat/0611014. (Cited on page 9.)

[33] S. Navin, *Diffraction in Pythia*, arXiv:1005.3894. (Cited on pages 11 and 244.)

[34] TOTEM collaboration, G. Antchev *et al.*, *Measurement of proton-proton elastic scattering and total cross-section at S**(1/2) = 7-TeV*, EPL **101** (2013) 21002. (Cited on page 10.)

[35] TOTEM collaboration, G. Antchev *et al.*, *Luminosity-independent measurements of total, elastic and inelastic cross-sections at* $\sqrt{s} = 7$ *TeV*, EPL **101** (2013) 21004. (Cited on page 10.)

[36] U. Amaldi, M. Jacob, and G. Matthiae, *Diffraction of Hadronic Waves*, Ann. Rev. Nucl. Part. Sci. **26** (1976) 385. (Cited on page 10.)

[37] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622. (Cited on pages 12 and 244.)

[38] NA50 collaboration, M. C. Abreu *et al.*, *Observation of a threshold effect in the anomalous J/ psi suppression*, Phys. Lett. **B450** (1999) 456. (Cited on page 14.)

[39] WA97 collaboration, E. Andersen *et al.*, *Strangeness enhancement at mid-rapidity in Pb Pb collisions at 158-A-GeV/c*, Phys. Lett. **B449** (1999) 401. (Cited on page 14.)

[40] STAR collaboration, K. H. Ackermann *et al.*, *Elliptic flow in Au + Au collisions at (S(NN))**(1/2) = 130 GeV*, Phys. Rev. Lett. **86** (2001) 402, arXiv:nucl-ex/0009011. (Cited on page 14.)

[41] ALICE collaboration, S. Acharya *et al.*, *Transverse momentum spectra and nuclear modification factors of charged particles in pp, p-Pb and Pb-Pb collisions at the LHC*, JHEP **11** (2018) 013, arXiv:1802.09145. (Cited on pages 14, 18, 19, 21, 97, 189, 191, 194, 200, 204, 208, 218, and 248.)

[42] CMS collaboration, V. Khachatryan *et al.*, *Charged-particle nuclear modification factors in PbPb and pPb collisions at* $\sqrt{s_{\mathrm{N\,N}}} = 5.02$ *TeV*, JHEP **04** (2017) 039, arXiv:1611.01664. (Cited on pages 14, 18, 189, 194, 200, 208, and 218.)

[43] ATLAS, G. Aad *et al.*, *Measurement of charged-particle spectra in Pb+Pb collisions at* $\sqrt{s_{\mathrm{NN}}} = 2.76$ *TeV with the ATLAS detector at the LHC*, JHEP **09** (2015) 050, arXiv:1504.04337. (Cited on page 14.)

[44] CMS collaboration, V. Khachatryan *et al.*, *Measurement of long-range near-side two-particle angular correlations in pp collisions at* $\sqrt{s}$ *=13 TeV*, Phys. Rev. Lett. **116** (2016) 172302, arXiv:1510.03068. (Cited on page 15.)

[45] CMS collaboration, S. Chatrchyan *et al.*, *Observation of Long-Range Near-Side Angular Correlations in Proton-Lead Collisions at the LHC*, Phys. Lett. **B718** (2013) 795, arXiv:1210.5482. (Cited on page 15.)

[46] J. L. Nagle and W. A. Zajc, *Small system collectivity in relativistic hadronic and nuclear collisions*, Ann. Rev. Nucl. Part. Sci. **68** (2018) 211, arXiv:1801.03477. (Cited on pages 15 and 208.)

[47] S. Schlichting and P. Tribedy, *Collectivity in Small Collision Systems: An Initial-State Perspective*, Adv. High Energy Phys. **2016** (2016) 8460349, arXiv:1611.00329. (Cited on page 15.)

[48] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250. (Cited on pages 16, 42, and 210.)

[49] R. Engel, D. Heck, and T. Pierog, *Extensive air showers and hadronic interactions at high energy*, Ann. Rev. Nucl. Part. Sci. **61** (2011) 467. (Cited on page 16.)

[50] Pierre Auger collaboration, A. Aab *et al.*, *The Pierre Auger Cosmic Ray Observatory*, Nucl. Instrum. Meth. **A798** (2015) 172, arXiv:1502.01323. (Cited on page 16.)

[51] A. Buckley *et al.*, *Rivet user manual*, Comput. Phys. Commun. **184** (2013) 2803, arXiv:1003.0694. (Cited on page 16.)

[52] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820. (Cited on pages 16, 42, and 210.)

[53] G. Ingelman and P. E. Schlein, *Jet Structure in High Mass Diffractive Scattering*, Phys. Lett. **B152** (1985) 256. (Cited on page 16.)

[54] T. Sjostrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, Phys. Rev. **D36** (1987) 2019. (Cited on page 16.)

[55] T. Sjostrand and P. Z. Skands, *Transverse-momentum-ordered showers and interleaved multiple interactions*, Eur. Phys. J. **C39** (2005) 129, arXiv:hep-ph/0408302. (Cited on pages 16 and 101.)

[56] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31. (Cited on page 17.)

[57] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047. (Cited on pages 17, 24, and 210.)

[58] T. Pierog *et al.*, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. **C92** (2015) 034906, arXiv:1306.0121. (Cited on pages 17, 24, 42, 54, 210, and 212.)

[59] T. Pierog and K. Werner, *EPOS Model and Ultra High Energy Cosmic Rays*, Nucl. Phys. Proc. Suppl. **B196** (2009) 102, arXiv:0905.1198. (Cited on page 17.)

[60] H. J. Drescher *et al.*, *Parton based Gribov-Regge theory*, Phys. Rept. **350** (2001) 93, arXiv:hep-ph/0007198. (Cited on page 17.)

[61] K. Werner, F.-M. Liu, and T. Pierog, *Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at RHIC*, Phys. Rev. **C74** (2006) 044902, arXiv:hep-ph/0506232. (Cited on page 17.)

[62] K. Werner, *Core-corona separation in ultra-relativistic heavy ion collisions*, Phys. Rev. Lett. **98** (2007) 152301, arXiv:0704.1270. (Cited on page 17.)

[63] C. A. Salgado *et al.*, *Proton-Nucleus Collisions at the LHC: Scientific Opportunities and Requirements*, J. Phys. **G39** (2012) 015010, arXiv:1105.3919. (Cited on pages 17 and 207.)

[64] J. L. Albacete *et al.*, *Predictions for cold nuclear matter effects in p+Pb collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV*, Nucl. Phys. **A972** (2018) 18, arXiv:1707.09973. (Cited on pages 17, 18, and 208.)

[65] N. Armesto, *Nuclear shadowing*, J. Phys. **G32** (2006) R367, arXiv:hep-ph/0604108. (Cited on pages 18 and 20.)

[66] V. Guzey, M. Strikman, and W. Vogelsang, *Observations on dA scattering at forward rapidities*, Phys. Lett. **B603** (2004) 173, arXiv:hep-ph/0407201. (Cited on pages 18, 195, 196, 201, and 202.)

[67] I. Helenius, K. J. Eskola, and H. Paukkunen, *Probing the small-x nuclear gluon distributions with isolated photons at forward rapidities in p+Pb collisions at the LHC*, JHEP **09** (2014) 138, arXiv:1406.1689. (Cited on pages 18, 196, 198, 199, 201, 215, 217, 218, and 248.)

[68] CMS collaboration, V. Khachatryan *et al.*, *Nuclear Effects on the Transverse Momentum Spectra of Charged Particles in pPb Collisions at* $\sqrt{s_{\rm NN}} = 5.02$ *TeV*, Eur. Phys. J. **C75** (2015) 237, arXiv:1502.05387. (Cited on pages 18, 19, 194, 195, and 248.)

[69] ATLAS collaboration, G. Aad *et al.*, *Transverse momentum, rapidity, and centrality dependence of inclusive charged-particle production in* $\sqrt{s_{NN}} = 5.02$ *TeV p + Pb collisions measured by the ATLAS experiment*, Phys. Lett. **B763** (2016) 313, arXiv:1605.06436. (Cited on pages 18, 19, 21, 189, 194, 195, and 208.)

[70] BRAHMS collaboration, I. Arsene *et al.*, *On the evolution of the nuclear modification factors with rapidity and centrality in d + Au collisions at s(NN)**(1/2) = 200-GeV*, Phys. Rev. Lett. **93** (2004) 242303, arXiv:nucl-ex/0403005. (Cited on pages 18, 19, 21, 196, 197, 201, 204, and 208.)

[71] PHENIX collaboration, C. Aidala *et al.*, *Nuclear-modification factor of charged hadrons at forward and backward rapidity in p+Al and p+Au collisions at* $\sqrt{s_{NN}} =$ 200 *GeV*, Phys. Rev. **C101** (2020) 034910, arXiv:1906.09928. (Cited on pages 18, 19, 22, 196, 197, 198, 201, 204, 208, 216, and 219.)

[72] K. J. Golec-Biernat and M. Wusthoff, *Saturation effects in deep inelastic scattering at low Q**2 and its implications on diffraction*, Phys. Rev. **D59** (1998) 014017, arXiv:hep-ph/9807513. (Cited on page 18.)

[73] H. Kowalski, T. Lappi, and R. Venugopalan, *Nuclear enhancement of universal dynamics of high parton densities*, Phys. Rev. Lett. **100** (2008) 022303, arXiv:0705.3047. (Cited on pages 18, 21, and 208.)

[74] ALICE collaboration, B. Abelev *et al.*, *Transverse momentum distribution and nuclear modification factor of charged particles in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Rev. Lett. **110** (2013) 082302, arXiv:1210.4520. (Cited on page 18.)

[75] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, *EPPS16: Nuclear parton distributions with LHC data*, Eur. Phys. J. **C77** (2017) 163, arXiv:1612.05741. (Cited on pages 20, 196, 208, 216, and 244.)

[76] K. Kovarik *et al.*, *nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework*, Phys. Rev. **D93** (2016) 085037, arXiv:1509.00792. (Cited on pages 20 and 208.)

[77] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, *Global analysis of nuclear parton distributions*, Phys. Rev. **D85** (2012) 074028, arXiv:1112.6324. (Cited on pages 20 and 208.)

[78] L. N. Lipatov, *Reggeization of the Vector Meson and the Vacuum Singularity in Nonabelian Gauge Theories*, Sov. J. Nucl. Phys. **23** (1976) 338. (Cited on page 21.)

[79] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, *The Pomeranchuk Singularity in Nonabelian Gauge Theories*, Sov. Phys. JETP **45** (1977) 199. (Cited on page 21.)

[80] I. I. Balitsky and L. N. Lipatov, *The Pomeranchuk Singularity in Quantum Chromodynamics*, Sov. J. Nucl. Phys. **28** (1978) 822. (Cited on page 21.)

[81] H1 collaboration, F. D. Aaron *et al.*, *A Precision Measurement of the Inclusive ep Scattering Cross Section at HERA*, Eur. Phys. J. **C64** (2009) 561, arXiv:0904.3513. (Cited on page 21.)

[82] L. V. Gribov, E. M. Levin, and M. G. Ryskin, *Semihard Processes in QCD*, Phys. Rept. **100** (1983) 1. (Cited on page 21.)

[83] A. H. Mueller and J.-w. Qiu, *Gluon Recombination and Shadowing at Small Values of x*, Nucl. Phys. **B268** (1986) 427. (Cited on page 21.)

[84] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, *The Color Glass Condensate*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 463, arXiv:1002.0333. (Cited on pages 21 and 244.)

[85] L. D. McLerran and R. Venugopalan, *Gluon distribution functions for very large nuclei at small transverse momentum*, Phys. Rev. **D49** (1994) 3352, arXiv:hep-ph/9311205. (Cited on pages 21 and 208.)

[86] F. Gelis, *Color Glass Condensate and Glasma*, Int. J. Mod. Phys. **A28** (2013) 1330001, arXiv:1211.3327. (Cited on page 21.)

[87] A. H. Rezaeian, *CGC predictions for p+A collisions at the LHC and signature of QCD saturation*, Phys. Lett. **B718** (2013) 1058, arXiv:1210.2385. (Cited on page 21.)

[88] J. L. Albacete, A. Dumitru, H. Fujii, and Y. Nara, *CGC predictions for p + Pb collisions at the LHC*, Nucl. Phys. **A897** (2013) 1, arXiv:1209.2001. (Cited on page 21.)

[89] T. Lappi and H. Mäntysaari, *Single inclusive particle production at high energy from HERA data to proton-nucleus collisions*, Phys. Rev. **D88** (2013) 114020, arXiv:1309.6963. (Cited on pages 21, 196, 208, and 216.)

[90] D. Antreasyan *et al.*, *Production of hadrons at large transverse momentum in 200-GeV, 300-GeV and 400-GeV p p and p n Collisions*, Phys. Rev. **D19** (1979) 764. (Cited on page 21.)

[91] PHENIX collaboration, S. S. Adler *et al.*, *Absence of suppression in particle production at large transverse momentum in S(NN)**(1/2) = 200-GeV d + Au collisions*, Phys. Rev. Lett. **91** (2003) 072303, arXiv:nucl-ex/0306021. (Cited on page 21.)

[92] PHOBOS collaboration, B. B. Back *et al.*, *Pseudorapidity dependence of charged hadron transverse momentum spectra in d+Au collisions at s(NN)**(1/2) = 200 GeV*, Phys. Rev. **C70** (2004) 061901, arXiv:nucl-ex/0406017. (Cited on pages 21 and 208.)

[93] ALICE collaboration, J. Adam *et al.*, *Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at* $\sqrt{s_{\rm NN}}$ *= 5.02 TeV*, Phys. Lett. **B760** (2016) 720, arXiv:1601.03658. (Cited on pages 22, 97, 98, 99, 102, 184, 204, and 212.)

[94] PHENIX collaboration, A. Adare *et al.*, *Spectra and ratios of identified particles in Au+Au and d+Au collisions at* $\sqrt{s_{NN}} = 200$ *GeV*, Phys. Rev. **C88** (2013) 024906, arXiv:1304.3410. (Cited on pages 22 and 204.)

[95] Z.-B. Kang, I. Vitev, and H. Xing, *Multiple scattering effects on inclusive particle production in the large-x regime*, Phys. Rev. **D88** (2013) 054010, arXiv:1307.3557. (Cited on pages 22, 196, 198, 204, 208, and 216.)

[96] A. Accardi, *Cronin effect in proton nucleus collisions: A Survey of theoretical models*, arXiv:hep-ph/0212148. (Cited on pages 22, 204, and 208.)

[97] Z.-B. Kang *et al.*, *Multiple scattering effects on heavy meson production in p+A collisions at backward rapidity*, Phys. Lett. **B740** (2015) 23, arXiv:1409.2494. (Cited on pages 22, 196, 198, 204, and 216.)

[98] ALICE collaboration, S. Acharya *et al.*, *Production of muons from heavy-flavour hadron decays in p-Pb collisions at* $\sqrt{s_{\rm NN}} = \mathbf{5.02}$ *TeV*, Phys. Lett. **B770** (2017) 459, arXiv:1702.01479. (Cited on page 22.)

[99] PHENIX collaboration, A. Adare *et al.*, *Cold-Nuclear-Matter Effects on Heavy-Quark Production at Forward and Backward Rapidity in d+Au Collisions at* $\sqrt{s_{NN}} = 200$ *GeV*, Phys. Rev. Lett. **112** (2014) 252301, arXiv:1310.1005. (Cited on page 22.)

[100] R. C. Hwa and C. B. Yang, *Final state interaction as the origin of the Cronin effect*, Phys. Rev. Lett. **93** (2004) 082302, arXiv:nucl-th/0403001. (Cited on page 22.)

[101] *Lhc minimum bias and underlying event working group conveners' meeting*, http://lpcc.web.cern.ch/sites/lpcc.web.cern.ch/files/wg_docs/minutes-16-03-2015-final.pdf. Accessed: 2020-01-07. (Cited on page 23.)

[102] ALICE collaboration, S. Acharya *et al.*, *The ALICE definition of primary particles*, Tech. Rep. ALICE-PUBLIC-2017-005, CERN, Jun, 2017. (Cited on pages 23 and 209.)

[103] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175. (Cited on page 24.)

[104] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001. (Cited on page 27.)

[105] E. Mobs, *The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019*, https://cds.cern.ch/record/2684277, July, 2019. General Photo. (Cited on pages 28 and 244.)

[106] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003. (Cited on page 29.)

[107] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004. (Cited on page 29.)

[108] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv:1207.7214. (Cited on page 29.)

[109] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv:1207.7235. (Cited on page 29.)

[110] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002. (Cited on page 29.)

[111] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769, Submitted to Nature Physics. (Cited on page 29.)

[112] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in charm decays*, Phys. Rev. Lett. **122** (2019) 211803, arXiv:1903.08726. (Cited on page 29.)

[113] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352. (Cited on pages 29, 31, 70, 209, and 244.)

[114] DELPHI collaboration, P. A. Aarnio *et al.*, *The DELPHI detector at LEP*, Nucl. Instrum. Meth. **A303** (1991) 233. (Cited on page 29.)

[115] LHCb collaboration, C. Elsässer, $b\bar{b}$ *production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html . (Cited on pages 30 and 244.)

[116] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fractions for* $B^0_{(s)} \to D_{(s)}\pi\pi\pi$ *and* $\Lambda^0_b \to \Lambda^+_c\pi\pi\pi$, Phys. Rev. **D84** (2011) 092001, Erratum ibid. **D85** (2012) 039904, arXiv:1109.6831. (Cited on page 31.)

[117] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of the* $\mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)$ *and* $\mathcal{B}(B^0 \to D^{*-}\mu^+\nu_\mu)$ *branching fractions using three-prong* τ*-lepton decays*, Phys. Rev. Lett. **120** (2018) 171802, arXiv:1708.08856. (Cited on page 31.)

[118] LHCb collaboration, R. Aaij *et al.*, *First evidence for the decay* $B^0_s \to \mu^+\mu^-$, Phys. Rev. Lett. **110** (2013) 021801, arXiv:1211.2674. (Cited on page 31.)

[119] LHCb collaboration, R. Aaij *et al.*, *Measurement of the relative rate of prompt* χ_{c0}, χ_{c1} *and* χ_{c2} *production at* $\sqrt{s} = 7\,TeV$, JHEP **10** (2013) 115, arXiv:1307.4285. (Cited on page 31.)

[120] LHCb collaboration, R. Aaij *et al.*, *Measurement of prompt-cross-section ratio* $\sigma(\chi_{c2})/\sigma(\chi_{c1})$ *in pPb collisions at* $\sqrt{s_{NN}} = 8.16\,TeV$, arXiv:2103.07349, to appear in Phys. Rev. C. (Cited on page 31.)

[121] LHCb collaboration, I. Bediaga *et al.*, *Framework TDR for the LHCb Upgrade: Technical Design Report*, Tech. Rep. CERN-LHCC-2012-007, CERN, Geneva, 2012. (Cited on page 32.)

[122] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808. (Cited on pages 33, 34, 35, and 244.)

[123] LHCb collaboration, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9** (2014) P12005, arXiv:1410.0149. (Cited on pages 33 and 49.)

[124] M. Alexander *et al.*, *Mapping the material in the LHCb vertex locator using secondary hadronic interactions*, JINST **13** (2018) P06008, arXiv:1803.07466. (Cited on page 33.)

[125] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, LHCb-PUB-2014-044. (Cited on page 35.)

[126] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893. (Cited on page 36.)

[127] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, arXiv:1708.00819. (Cited on page 36.)

[128] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, tech. rep., CERN, Geneva, Jan, 2011. LHCb. (Cited on page 38.)

[129] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, tech. rep., CERN, Geneva, May, 2007. (Cited on page 38.)

[130] M. Needham, *Performance of the Track Matching*, tech. rep., CERN, Geneva, Oct, 2007. (Cited on page 38.)

[131] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444. (Cited on page 38.)

[132] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251. (Cited on pages 39, 107, 162, 211, 224, and 249.)

[133] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2019) 1, arXiv:1803.00824. (Cited on page 39.)

[134] LHCb collaboration, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, JINST **14** (2019) P11023, arXiv:1909.02957. (Cited on page 39.)

[135] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759. (Cited on page 39.)

[136] R. Aaij *et al.*, , LHCb-DP-2020-001, in preparation. (Cited on page 40.)

[137] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055. (Cited on page 40.)

[138] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249. (Cited on page 41.)

[139] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023. (Cited on pages 42, 43, and 210.)

[140] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152. (Cited on pages 42 and 210.)

[141] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv:hep-ph/0506026. (Cited on page 42.)

[142] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270. (Cited on pages 42 and 210.)

[143] R. Brun *et al.*, *root-project/root: v6.18/02*, Aug., 2019. doi: 10.5281/zenodo.3895860. (Cited on page 43.)

[144] *Stripping project webpage*, http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/. Accessed: 2018-12-20. (Cited on page 48.)

[145] Y. Gao *et al.*, *Study of J/ψ production and cold nuclear matter effects in proton-lead collisions at $\sqrt{s_{\rm NN}} = 5$ TeV*, LHCb-ANA-2013-059. (Cited on page 48.)

[146] E. A. Maurice *et al.*, *Study of Cold Nuclear Matter effect with prompt D^0 meson production in p*Pb *collisions at LHCb.*, LHCb-ANA-2016-012. (Cited on page 48.)

[147] Y. Gao *et al.*, *Prompt Λ_c^+ production in p*Pb *collisions at $\sqrt{s_{\rm NN}} = 5.02\,$TeV* , LHCb-ANA-2018-014. (Cited on page 48.)

[148] https://twiki.cern.ch/twiki/bin/viewauth/LHCbPhysics/LHCbLuminosity. Accessed: 2021-03-02. (Cited on page 50.)

[149] V. Balagura. Private communication. Date: 2021-03-04. (Cited on pages 51 and 52.)

[150] M. Schmelling, *Luminosity calibration for the leading bunch nobias-triggered data from the 2015 early measurement runs*, Tech. Rep. LHCb-INT-2017-015. CERN-LHCb-INT-2017-015, CERN, Geneva, Jun, 2017. (Cited on pages 51 and 52.)

[151] *Jira task: Improved beam-option for 2013 ppb/pbp data*, https://its.cern.ch/jira/browse/LHCBGAUSS-1375. Accessed: 2018-02-05. (Cited on page 54.)

[152] *Jira task: V0 ppb epos 2013 mc issue*, https://its.cern.ch/jira/browse/LHCBGAUSS-1493. Accessed: 2018-12-20. (Cited on page 54.)

[153] X.-N. Wang and M. Gyulassy, *HIJING: A Monte Carlo model for multiple jet production in p p, p A and A A collisions*, Phys. Rev. **D44** (1991) 3501. (Cited on page 54.)

[154] W.-T. Deng, X.-N. Wang, and R. Xu, *Hadron production in p+p, p+Pb, and Pb+Pb collisions with the HIJING 2.0 model at energies available at the CERN Large Hadron Collider*, Phys. Rev. **C83** (2011) 014915, arXiv:1008.1841. (Cited on page 54.)

[155] *Rope hadronization in pythia 8.2*, https://pythia.org/manuals/pythia8245/Welcome.html. Accessed: 2020-10-10. (Cited on pages 56 and 101.)

[156] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762** (2016) 012036, arXiv:1608.05806. (Cited on page 64.)

[157] A. Rogozhnikov, *hep_ml python library*, https://arogozhnikov.github.io/hep_ml/reweight.html. Accessed: 2019-10-04. (Cited on page 64.)

[158] https://twiki.cern.ch/twiki/bin/view/LHCbInternal/TrackingEffRatio. Accessed: 2020-04-13. (Cited on page 68.)

[159] https://twiki.cern.ch/twiki/bin/view/LHCb/TrackingEffAbsLength. Accessed: 2019-05-07. (Cited on page 72.)

[160] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, LHCb-PUB-2017-011. (Cited on pages 75 and 211.)

[161] M. Needham, *Clone track identification using the Kullback-Liebler distance*, CERN-LHCB-2008-002, LPHE-2008-02. (Cited on page 75.)

[162] *Lhcb software doxygen documentation, lhcb::mcvertex class reference*, https://lhcb-doxygen.web.cern.ch/lhcb-doxygen/gauss/v54r5/dd/d97/class_l_h_cb_1_1_m_c_vertex.html#a2794a8fc7a81be4844ad8b4a61aed17e. Accessed: 2021-03-11. (Cited on page 77.)

[163] ALICE collaboration, J. Adam *et al.*, *Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions*, Nature Phys. **13** (2017) 535, arXiv:1606.07424. (Cited on pages 97, 98, 99, and 212.)

[164] *Tspline3 class reference, root reference guide*, https://root.cern.ch/doc/master/classTSpline3.html. Accessed: 2020-10-23. (Cited on pages 97 and 189.)

[165] ALICE collaboration, B. B. Abelev *et al.*, *Production of* $\Sigma(1385)^{\pm}$ *and* $\Xi(1530)^{0}$ *in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. **C75** (2015) 1, arXiv:1406.3206. (Cited on page 97.)

[166] LHCb collaboration, R. Aaij *et al.*, *Measurement of prompt hadron production ratios in pp collisions at* $\sqrt{s} = 0.9$ *and 7 TeV*, Eur. Phys. J. **C72** (2012) 2168, arXiv:1206.5160. (Cited on pages 97, 100, and 212.)

[167] C. Bierlich, *Microscopic collectivity: The ridge and strangeness enhancement from string–string interactions*, Nucl. Phys. **A982** (2019) 499, arXiv:1807.05271. (Cited on pages 97 and 212.)

[168] ALICE collaboration, J. Adam *et al.*, *Multi-strange baryon production in p-Pb collisions at* $\sqrt{s_{\mathbf{NN}}} = 5.02$ *TeV*, Phys. Lett. **B758** (2016) 389, arXiv:1512.07227. (Cited on pages 101, 102, and 212.)

[169] ALICE collaboration, B. B. Abelev *et al.*, *Multiplicity Dependence of Pion, Kaon, Proton and Lambda Production in p-Pb Collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Lett. **B728** (2014) 25, arXiv:1307.6796. (Cited on pages 101, 102, and 212.)

[170] https://twiki.cern.ch/twiki/bin/view/LHCbInternal/LHCbTrackingEfficiencies. Accessed: 2020-04-13. (Cited on pages 108, 110, and 245.)

[171] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001. (Cited on pages 116 and 117.)

[172] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083. (Cited on pages 120, 233, 234, and 235.)

[173] T. Adye, *Unfolding algorithms and tests using RooUnfold*, in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, (Geneva), pp. 313–318, 2011, arXiv:1105.1160. doi: 10.5170/CERN-2011-006.313. (Cited on page 155.)

[174] *Roounfold documentation webpage*, https://http://hepunx.rl.ac.uk/webtempfiles/adye/public_html/software/unfold/RooUnfold.html. Accessed: 2019-03-15. (Cited on page 155.)

[175] https://svnsrv.desy.de/websvn/wsvn/General.unfolding/RooUnfold/trunk/. Accessed: 2019-05-13. (Cited on page 155.)

[176] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Meth. **A362** (1995) 487. (Cited on page 155.)

[177] C. Loizides, J. Kamin, and D. d'Enterria, *Improved Monte Carlo Glauber predictions at present and future nuclear colliders*, Phys. Rev. **C97** (2018) 054910, arXiv:1710.07098, [Erratum: Phys. Rev. C99, 019901 (2019)]. (Cited on pages 188 and 215.)

[178] LHCb collaboration, R. Aaij *et al.*, *Measurement of prompt charged-particle production in pp collisions at* $\sqrt{s} = 13$ *TeV*, LHCb-PAPER-2021-010, in preparation. (Cited on pages 189, 190, 203, 215, and 248.)

[179] ALICE collaboration, J. Adam *et al.*, *Pseudorapidity and transverse-momentum distributions of charged particles in proton–proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. **B753** (2016) 319, arXiv:1509.08734. (Cited on page 189.)

[180] CMS, TOTEM collaborations, S. Chatrchyan *et al.*, *Measurement of pseudorapidity distributions of charged particles in proton-proton collisions at $\sqrt{s} = 8$ TeV by the CMS and TOTEM experiments*, Eur. Phys. J. **C74** (2014) 3053, arXiv:1405.0722. (Cited on pages 189, 191, and 248.)

[181] J. W. Cronin *et al.*, *Production of hadrons with large transverse momentum at 200, 300, and 400 GeV*, Phys. Rev. **D11** (1975) 3105. (Cited on pages 193, 208, and 215.)

[182] ALICE, J. Adam *et al.*, *Centrality dependence of particle production in p-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV*, Phys. Rev. **C91** (2015) 064905, arXiv:1412.6828. (Cited on page 194.)

[183] L. L. Frankfurt and M. I. Strikman, *High-Energy Phenomena, Short Range Nuclear Structure and QCD*, Phys. Rept. **76** (1981) 215. (Cited on page 195.)

[184] L. L. Frankfurt and M. I. Strikman, *Hard Nuclear Processes and Microscopic Nuclear Structure*, Phys. Rept. **160** (1988) 235. (Cited on page 195.)

[185] S. Dulat *et al.*, *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. **D93** (2016) 033006, arXiv:1506.07443. (Cited on pages 196 and 216.)

[186] O. Boente García, *Recent results from heavy ion collisions at lhcb*, URL: https://indico.bnl.gov/event/9726/contributions/45494/attachments/33676/54197/DIS2021_oboente.pdf. presented on 13/4/2021. (Cited on page 203.)

[187] ALICE collaboration, S. Acharya *et al.*, *Nuclear modification factor of light neutral-meson spectra up to high transverse momentum in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV*, arXiv:2104.03116. (Cited on pages 204, 208, and 219.)

[188] A. B. Kaidalov and K. A. Ter-Martirosyan, *Multihadron production at high energies in the model of quark gluon strings*, Sov. J. Nucl. Phys. **40** (1984) 135. (Cited on page 207.)

[189] A. Capella, U. Sukhatme, C.-I. Tan, and J. Tran Thanh Van, *Dual parton model*, Physics Reports **236** (1994) 225. (Cited on page 207.)

[190] D. d'Enterria *et al.*, *Constraints from the first LHC data on hadronic event generators for ultra-high energy cosmic-ray physics*, Astropart. Phys. **35** (2011) 98, arXiv:1101.5596. (Cited on page 207.)

[191] F. Arleo, F. Cougoulic, and S. Peigné, *Fully coherent energy loss effects on light hadron production in pA collisions*, JHEP **09** (2020) 190, arXiv:2003.06337. (Cited on page 208.)

[192] T. Sjöstrand *et al.*, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012. (Cited on page 212.)